1

OPTICAL DIGITAL IMAGE STORAGE SYSTEM

Project Report

MARCH 1991

National Archives and Records Administration

Archival Research and Evaluation Staff

National Archives

Washington, DC 20408

March 18, 1991

Over the past decade, we have seen a tremendous increase in the use of computer-based systems to create, maintain, and access records of the federal government. Commensurate with this has been the response of the producers and developers of information systems equipment and software to provide new methods of storing and manipulating electronic records. One of these technologies is the capture and use of digital images or "electronic photographs" of documents. Since digital images comprise extraordinarily large volumes of digital data, another development drawing significant interest is the evolution of optical storage media that can provide compact storage for the massive capacities required by digital imaging systems.

Government agencies and private sector enterprises have been quick to recognize the potential of systems exploiting these technologies to manage large numbers of records effectively and efficiently where indexed access is a prime consideration.

In February 1984, the National Archives undertook the Optical Digital Image Storage System project, a research pilot to test and evaluate the feasibility, costs, and benefits of using digital imaging technology in support of archival programs. In the following five years, a team from our technology assessment unit, the Archival Research and Evaluation Staff, and our major records custodial office, the Office of the National Archives, developed specifications for, procured, and ran a large-scale pilot production facility that was used to capture and test a representative sample comprising a quarter million documents from the Civil War era.

This report documents all project activities over the five-year period including preparatory work leading to the undertaking of the pilot and details of the actual system operations. A number of analyses are presented comparing the prospective use of such a system compared to the current methods employed by the National Archives. For those readers unfamiliar with digital imaging or optical disk technologies, the report includes a monograph which provides a basic introduction to them.

By releasing this report to the public, the National Archives invites an exchange of views with the archival community and related professions on the implications of these technologies. We look forward to your comments on the report and the opportunity to further explore the potential of their application in archival administration.

DON W. WILSON
Archivist of the United States

Project Director:

William L. Hooton

Primary Project Staff:

Barry S. Roginski

Michael A. Goldman

ARCHIVAL RESEARCH AND EVALUATION STAFF

Director:	William M. Holmes, Jr.
Assistant Director:	Charles M. Dollar
Staff Contributors:	Beverly S. Hacker Avra Michelson Thomas E. Weir

Acknowledgements:

The Archival Research and Evaluation Staff would like to thank Trudy H. Peterson and Michael J. Kurtz, Assistant Archivist and Deputy Assistant Archivist respectively, of the Office of the National Archives for providing special assistance and support over the life of the project. Without their contributions, along with those of many of their staff, the successful completion of the project would not have been possible. We would also like to extend our appreciation to the staff of the Office of Management and Administration who also contributed to the success of the project.

The following is a partial alphabetical list of those who directly contributed to ODISS.

David R. Allder	*Irene C. Anthony*	*Anna M. Barnaba*
Juanita M. Bennett	*Carolyn L. Bernaski*	*Kevin L. Bradley*
Jean Bray	*Thomas E. Brown*	*Tod J. Butler*
Alan R. Calmes	*Gwen Courtney*	*Richard F. Cox, Jr.*
Wanda Y. Curry	*William H. Davis*	*Andrew L. Dyer*
Sharon K. Fawcett	*Mary M. Fingers*	*Cynthia G. Fox*
Monroe Freeman	*Barbara J. P. Frye*	*Teresa J. Gantt*
Michael E. Getsey	*Cynthia Ghee*	*Edwin Gleaves*
Roger J. Gorg	*Geraldine M. Green*	*Stephen Hannestad*
Kenneth E. Harris	*Francis J. Heppner*	*Brenda A. Kepley*
Mary D. Kotner	*John F. Kreinheder*	*Maida H. Loescher*
Elaine McKoy	*Jerry F. Misko*	*Geraldine N. Phillips*
Constance Potter	*Sandra C. Powell*	*Eva Scott-Cora*
Vanessa Skaggs	*Barbara L. Smith*	*Joan G. Smith*
Michael J. Stanchie	*Will Templeton*	*T. Wayne Tracey*
Jimma D. Tufa	*Bobbye C. West*	*James Whittington*

OPTICAL DIGITAL IMAGE STORAGE SYSTEM PROJECT REPORT

TABLE OF CONTENTS

PREFACE xii

1 ARCHIVAL MANAGEMENT AND TECHNOLOGY SUMMARY 2
1.1 Project Origin 2
1.1.1 Goals 2
1.1.2 Test Sample 2
1.2 Technology Summary 3
1.2.1 Digital Imaging 3
1.2.2 Optical Disk 4
1.2.3 Computer Retrieval 4
1.3 Archives and Management Issues 4
1.3.1 Background 4
1.3.2 Document Conversion Issues 5
1.3.2.1 Document Preparation 5
1.3.2.2 Image Capture 6
1.3.2.3 Image Utility 8
1.3.2.4 Image Stability 12
1.3.3 Document Retrieval Issues 14
1.3.3.1 Document Access 14
1.3.3.1.1 Speed and Relevance 15
1.3.3.1.2 Simplicity of User System Interface 17
1.3.3.1.3 Enhanced Retrieval Capability 18
1.3.3.1.4 Decentralized Distribution 18
1.4 Cost Effectiveness 19
1.4.1 Document Throughput 19
1.4.2 Space Reduction 21
1.4.3 Improved Access 21
1.4.4 Cost-Benefit Concerns 22

2 PROJECT HISTORY AND PURPOSE 26
2.1 Origins of the ODISS Project 26
2.2 Project Objectives and Procedures 26
2.3 ODISS Design and Technical Requirements 28
2.4 System Acquisition and Implementation Process 29

3 EXISTING NARA PROCESSES AND TECHNOLOGY UTILIZATION 36
3.1 Paper Records 36
3.1.1 Physical Characteristics 36
3.1.2 Administration of Permanent Records 37
3.1.3 Document Preservation and Conservation 38
3.1.4 Retrieval And Finding Aids 39
3.2 NARA Micrographics Policy and Operations 39
3.2.1 Evolution of Micrographics in the National Archives 39
3.2.2 Role in Records Storage and Preservation 40
3.2.3 Administrative Management 40
3.2.4 System Operations 41

3.2.4.1 Camera Area Operations and Production Statistics 41
3.2.4.2 Film Processing Operations and Production Statistics 42
3.2.4.3 Quality Control Operations 42
3.2.4.4 Testing and Storage Requirements 43
3.2.4.5 Duplication Operations 44
3.2.4.6 Production Problems 44
3.2.4.7 Document Handling Considerations During Conversions 45
3.2.5 Information Retrieval from Microforms 45
3.2.5.1 Utilization for Research 46
3.2.5.2 Image Quality Considerations 47
3.3 CMSR Reference 47
3.3.1 Reference Activity 47
3.3.1.1 Staff and Organization 49
3.3.1.2 Walk-in Public Reference 49

4 ODISS SUBSYSTEM DESCRIPTIONS 54
4.1 General System Concept 54
4.1.1 File Data Structure 54
4.1.2 Conversion 54
4.1.3 Storage 55
4.1.4 Retrieval 55
4.1.5 Duties of the ODISS System Manager 55
4.2 Hardware and Software Configuration 56
4.2.1 Major Subsystems 56
4.2.2 Digital Image Scanners 57
4.2.2.1 High Speed Scanner 57
4.2.2.2 Low Speed Paper Scanners 57
4.2.2.2.1 Binary Scanner 58
4.2.2.2.2 Gray Scale Scanner 58
4.2.2.3 Multi-Format Microform Scanner 58
4.2.3 Workstation Subsystem 59
4.2.3.1 Indexing 59
4.2.3.2 Quality Control 59
4.2.3.3 Rescanning and Replacement 60
4.2.3.4 Retrieval 61
4.2.3.4.1 Staff Retrieval 61
4.2.3.4.2 Public Retrieval 61
4.2.3.4.3 Remote Retrieval 62
4.2.4 Archive Subsystem 62
4.2.5 System Manager, and Initiate and Monitor Subsystem 62
4.2.5.1 System Manager Terminal 62
4.2.5.2 CSE/ARS Terminal 63
4.2.5.3 IMS/Archive Control Terminal 63

5 ODISS TEST PLAN DESCRIPTION 66
5.1 Testing Goals 66
5.2 Test Sample Selection 66
5.3 Test Sample Attributes 66
5.3.1 CMSR Documents 66
5.3.2 Non-CMSR Documents 68
5.3.3 Microform Samples 68

5.4 Testing Facilities and Locations 68
5.5 Test Duration 68
5.6 Constraints and Considerations 69
5.7 Measurement of User Satisfaction 69
5.8 Data Collection and Analysis Methodology 69
5.8.1 Test Criteria Framework 69
5.8.2 Test Criteria Descriptions 70
5.8.2.1 High Speed Scanning 70
5.8.2.2 Image Quality 72
5.8.2.3 Production Workflow 74
5.8.2.4 Indexing 75
5.8.2.5 Quality Control 77
5.8.2.6 Low Speed Scanning and Enhancement 78
5.8.2.7 System Manager 79
5.8.2.8 System Operations 80
5.8.2.9 Microform Scanning 81
5.8.2.10 Index Storage 81
5.8.2.11 Image Storage 82
5.8.2.12 On-Site Reference 83
5.8.2.13 Remote Reference 84
5.8.2.14 Hardcopy Output 85

6 PROJECT OPERATIONS ANALYSIS AND TEST RESULTS 88
6.1 Document Preparation For The ODISS Project 88
6.1.1 Tennessee CMSR Records 88
6.1.2 Differences Between Preparation for ODISS and Microfilming 90
6.1.3 Lessons Learned 90
6.2 High Speed Scanning 91
6.2.1 Ease of Use of the Workstation 91
6.2.2 Production Rate and Throughput 91
6.2.2.1 CMSR Sample 92
6.2.2.2 File Control Considerations 92
6.2.2.3 Handling Image Anomalies 93
6.2.2.4 Pension and Bounty Land Warrant Sample 94
6.2.2.5 Government Printing Office Sample 94
6.2.3 Scanner Transport Considerations 94
6.2.3.1 Dealing with Different Document Characteristics 94
6.2.3.2 Use of Polyester Sleeves 95
6.2.3.3 Color Sensitivity 95
6.2.3.4 Sensor Placement 95
6.2.3.5 Other Considerations 96
6.2.4 Suggested Improvements 96
6.3 Indexing 98
6.3.1 Ease Of Learning the Workstation 98
6.3.2 Operators' Views 100
6.3.3 Production and Throughput Rates 100
6.3.4 Analysis Of Data 102
6.4 Quality Control 103
6.4.1 Ease of Learning the Workstation 103
6.4.2 Ease of Use of the Workstation 103
6.4.3 Operators' Views 104

6.4.4 Production and Throughput Rates 105
6.4.5 Image Quality Rejection Rate 107
6.4.6 Analysis of Data 108
6.5 Low Speed Scanning and Image Enhancement 108
6.5.1 Gray Scale Image Enhancement Workstation 108
6.5.2 Binary Scanner 111
6.5.3 IPT Scan Optimizer 111
6.5.4 Production Statistics 112
6.5.5 Testing of the Workstation 115
6.5.6 Pension and Bounty Land Warrant Sample 117
6.5.7 Government Printing Office Sample 119
6.6 Multiformat Microform Scanner 120
6.6.1 Operability and Ease of Use of the Workstation 120
6.6.2 Staff Comparisons of Image Quality from Scans of Paper and Film 121
6.6.2.1 CMSR Tennessee Infantry Document Tests 122
6.6.2.2 Government Printing Office Document Tests 124
6.7 Optical Storage and Archiving 125
6.7.1 Archive Process Overview 125
6.7.2 Archives Workstation 126
6.7.3 Optical Disk Security Backup System 126
6.7.4 Optical Disk Longevity 133
6.7.5 Analysis of WORM Disk Capacity 133
6.7.6 Operational Experiences 134
6.8 Staff Retrieval 135
6.8.1 Test Design and Procedures 135
6.8.2 Test Implementation 135
6.8.3 Ease of Learning and Use of the Workstation 136
6.8.4 Production Rate 137
6.8.5 Search Accuracy 137
6.8.6 Analysis of Test Data 138
6.9 Public Retrieval 139
6.9.1 Test Procedures 140
6.9.2 Test Results 140
6.9.3 Analysis of the Test Data 142
6.9.4 Public Survey 142
6.9.5 Analysis of the Survey Data 144
6.10 System Manager 144
6.10.1 Ease of Use 144
6.10.2 Specific Problems and Suggestions for Improvement 146
6.10.3 Overall Effectiveness 148
6.11 Remote Workstation 148
6.11.1 Workstation Configuration 148
6.11.2 Operational Experiences 148
6.12 Production and Evaluation of Image Quality 149
6.12.1 Technical and Subjective Considerations 150
6.12.2 Image Enhancement Issues 150
6.12.3 Photographic and Electronic Imaging 151
6.13 General Testing Issues and Results 152
6.13.1 Validity of the Original Design Concept 152
6.13.2 Modifications to System Operations and Workflow 153
6.13.3 System Modeling 154

6.13.3.1 Operational Use of the Modeling Software 154
6.13.3.2 Analysis and Findings 155
6.13.3.3 Optimum Workstation Configuration 156
6.13.3.4 Optimum Performance Potential 160
6.13.4 System Maintenance 160
6.13.5 Personnel and Staffing 161
6.13.5.1 Training and Operations 161
6.13.5.2 Cross-training 161
6.13.5.3 Operator Performance 161
6.13.6 Ergonomic Factors 162

A. OVERVIEW OF DIGITAL IMAGE AND OPTICAL MEDIA TECHNOLOGIES 164
A.1 Digital Image Technology 164
A.1.1 Introduction 164
A.1.2 Document Conversion 164
A.1.2.1 What is a Digital Image? 164
A.1.2.2 Scanning From Different Sources 165
A.1.2.3 Scanners 165
A.1.2.4 Image Enhancement 166
A.1.2.5 File Management and Control 169
A.1.2.6 Indexing 169
A.1.2.7 Quality Control and Assurance 171
A.1.2.8 Data Compression and File Size 171
A.1.2.9 Image [Input] Data Buffer Storage 172
A.1.3 Image Retrieval 173
A.1.3.1 Locating the Image File 173
A.1.3.2 Image [Output] Data Buffer Storage 173
A.1.3.3 Image Workstations 174
A.1.3.3.1 Display Density 174
A.1.3.3.2 Simultaneous Display of Image and Character Data 174
A.1.3.3.3 Display Features 175
A.1.3.4 Printers 178
A.1.3.4.1 Print Density 178
A.1.3.4.2 Print Speed 179
A.2 Optical Media Technology 179
A.2.1 Introduction 179
A.2.2 What is an Optical Disk? 179
A.2.3 Optical Storage Formats 189
A.2.3.1 Analog Videodiscs 189
A.2.3.2 Write-Once Digital Optical Disks 193
A.2.3.3 Rewritable Digital Optical Disks 193
A.2.3.4 Digital Read-Only Optical Disks 195
A.2.3.5 Digital Optical Tape 195
A.2.3.6 Digital Optical Cards 195
A.2.4 Write-Once Disk Recording Methodologies 197
A.2.5 Write-Once Recording Strategies 201
A.2.6 Automated Retrieval Devices 201
A.2.7 Optical Media Longevity and Stability 202
A.2.8 Legality of Digital Images From Optical Disks 202
A.2.9 Standardization 203

B. DETAILED ODISS SUBSYSTEM DESCRIPTIONS 206
B.1 Basic System Concept 206
B.1.1 Configurational Scheme 206
B.1.2 Capture and Retrieval Process 206
B.2 Digital Image Scanning 210
B.2.1 High Speed Paper Scanner 210
B.2.2 Low Speed Paper Scanners 222
B.2.3 Multiformat Microform Scanner 230
B.3 Image Enhancement and Quality Analysis 231
B.4 Production Throughput Capabilities 235
B.5 Indexing 236
B.5.1 Subject Terms For CMSR Files 236
B.5.2 Station Workflow 236
B.5.3 Hardware Configuration 237
B.5.4 Software Capabilities 237
B.6 Quality Control 239
B.6.1 Purposes 239
B.6.2 Station Workflow 239
B.6.3 Hardware Configuration And Software Capabilities 240
B.7 Digital Storage 240
B.7.1 In-Process Image and Index Data 241
B.7.2 Optical Disk Archival Storage 241
B.7.3 Archives Subsystem 245
B.8 Staff Retrieval 246
B.8.1 Station Workflow 246
B.8.2 Hardware Configuration and Software Capabilities 250
B.8.3 Non-CMSR Files 250
B.9 Public Retrieval 250
B.9.1 Station Workflow 252
B.9.2 Hardware Configuration And Software Capabilities 260
B.9.3 Adequacy of Screen Instructions 260
B.10 Remote Workstation 260
B.10.1 Configuration 260
B.10.2 Operation 260
B.10.3 Hardware 262
B.11 Laser Printer Subsystem 262
B.11.1 Hardware Configuration and Software Capabilities 264
B.11.2 Operation 264
B.12 System Manager 266
B.12.1 Hardware Configuration 266
B.12.2 Operations 267
B.12.3 System Manager Duties 278

C. COMPILED SYSTEM PERFORMANCE DATA 282
C.1 Tennessee CMSR File Sample 282
C.1.1 Quantity Converted 282
C.1.2 File Size 282
C.2 Conversion Statistics 283
C.2.1 High Speed Scanner Totals 283
C.2.2 Indexing Totals 284
C.2.3 Quality Control Totals 284

C.2.4 Low Speed Scanner Totals 285
C.3 Analysis of the Conversion Statistics 285
C.3.1 Optical Disk Storage Capacity 285
C.3.2 High Speed Scanning 286
C.3.3 Image Quality Rejection Rate 286
C.3.4 Average Daily Production Rates 287
C.3.4.1 Rates Based on the Full Available Work Period 287
C.3.4.2 Rates Based on the Days with Management Reports 287

D. COST ANALYSIS 290
D.1 Cost Analysis Methodology 290
D.1.1 Generic Application Description 290
D.1.2 Data Acquisition 292
D.1.3 Assumptions and Constraints 292
D.2 Existing Paper Records System 294
D.2.1 Description 294
D.2.2 Derivation of Costs 294
Advantagos and Disadvantages 295
D.3 Manual Microfilm System 297
D.3.1 Description 297
D.3.2 Derivation of Costs 297
Advantages and Disadvantages 299
D.4 Upgraded (CAR) Microfilm System 301
D.4.1 Description 301
D.4.3 Advantages and Disadvantages 305
D.5 Digital Image/Optical Disk System 307
D.5.1 Description 307
D.5.2 Derivation of Costs 308
D.5.3 Advantages and Disadvantages 309
D.6 Upgraded (CAR) Microfilm System Using Service Bureau Conversion 311
D.6.1 Description 311
D.6.2 Derivation of Costs 311
D.6.3 Advantages and Disadvantages 312
D.7 Digital Image/Optical Disk System Using Service Bureau Conversion 315
D.7.1 Description 315
D.7.2 Derivation of Costs 315
D.7.3 Advantages and Disadvantages 316
D.8 Comparison of Costs of System Alternatives 319
D.9 Interpretation of the Model 322

E. DATA COLLECTION FORMS 324

F. RESEARCH TEST IMPLEMENTATION CONSIDERATIONS 340
F.1 Rationale for Research Testing 340
F.2 Role of the System Integrator 341
F.3 ODISS System Design Review Process 341
F.4 Factory Acceptance and On-Site Testing 342
F.5 ODISS Facility Design and Construction 343
F.5.1 Computer Room Environment 343
F.5.2 Fire Safety and Control Systems 344
F.5.3 Electrical and Signal Cable Installation 344

F.6 Equipment Floorplan Design 345
F.7 Ergonomic Workstation Furniture Specification 345
F.8 Production Staff and User Training 349
F.9 System Documentation 349

G. NARA MICROGRAPHICS PROGRAM 352
G.1 Technology Overview 352
G.2 Camera Area Equipment and Operations 353
G.2.1 Equipment 353
G.2.2 Staffing 354
G.2.3 Production Costs 354
G.3 Processing Equipment and Operations 355
G.3.1 Equipment 355
G.3.2 Staffing 356
G.4 Quality Control Equipment and Operations 356
G.4.1 Equipment 356
G.4.2 Staffing 356
G.5 Duplication Equipment and Operations 356
G.5.1 Equipment 356
G.5.2 Staffing 356
G.5.3 Production Costs 357
G.6 Future Plans 357

H. PHOTOGRAPHS OF ODISS EQUIPMENT 360

I. GLOSSARY OF TERMS 370

LIST OF FIGURES

Microfilm Research Process 48
CMSR Mail-in Search Process 50
CMSR Walk-in Service 51
Scanner Sensor Placement 97
Digital Image Enhancement 110
Unreadable Document 113
IPT Enhanced Image 114
Archive Subsystem Screen: Menu 127
Archive Subsystem Screen: Ready to Archive 128
Archive Subsystem Screen: Initiate Archive Process 129
Archive Subsystem Screen: Initiate Archive? (Y/N) 130
Archive Subsystem Screen: Progress Monitoring 131
Archive Subsystem Screen: Available Disk Space 132
Total Operational Time per File 158
Optimal Workstation Configuration 159
Histogram 168
Image and Character Terminals 176
Image Terminal 177
Optical Block 181
Optical Block Mechanism 182
Beam Grate 183
Principle of PBS 184
Role of PBS 185
Principle of Tracking 186
Principle of Focus Servo 187
Principle of Focusing 188
Write-Once Disk 190
Videodisc 191
Videodisc Production Sequence 192
Magneto Optic Recording Principle 194
Magneto Optic Reading Principle 196
Write-Once Disk Writing Methods 198
Reflection of a Laser Beam 199
Diffraction of a Laser Beam 200
System Block Diagram 207
Capture and Storage Subsystems 208
Retrieval Operation 209
Docuscan DS-4000 High Speed Transport 212
Scanner Button/Indicator Panel 216
Operational Control Terminal; Awaiting Block Open 218
Operational Control Terminal; Ready for Scanning 219
Operational Control Terminal; Next File Ready 220
Operational Control Terminal; Block Closed 221
Mode Menu 224
Main Options Menu 227
Workstation Subsystem 238
Capture Subsystem 242
Archive Subsystem 244
CMSR Search Screen 247

Highlighted File 248
Displayed Image 249
Image Search Results 251
Upside Down Image 253
180 Degree Rotation 254
Sideways Image 255
90 Degree Rotation 256
150 DPI Image 257
200 DPI Image 258
Zoom Mode 259
Remote Link 261
Remote Search Screen 263
Printer Subsystem 265
Cumulative Cost Graph 321
Indexing Input Data Collection Form 324
Standard Procedures For Indexing 327
Quality Control Data Collection Form 329
CMSR Search Batch 331
CMSR File Search Form 332
Staff Reference Data Collection Form 334
Public Reference Workstation Data Collection Form 337
ODISS Floor Plan 346
Workstation Designs 348
High Speed Scanner 360
Low Speed Scanner Station 361
Microfilm Scanner Station 362
Index and Quality Control Stations 363
Optical Disk Jukebox 364
Retrieval and Printing Stations 365
System Manager Station 366
Halon Fire Control Panel 367

♦ Some figures courtesy of Sony Corporation

LIST OF TABLES

Indexing Workstation - Ease of Learning 99
Indexing Workstation - Ease of Use 99
Indexing Wait Times - November 1988 102
Indexing Wait Times - December 1988 102
Quality Control Workstation - Ease of Learning 103
Quality Control Workstation - Ease of Use 104
Quality Control Workstation - Functional Evaluation 104
Total Production At Quality Control 105
Quality Control Production Rates - All Work Days 106
Quality Control Production Rates - Active Work Days 106
Quality Control Timings - December 1988 and January 1989 107
Quality Control Timings - February 1989 107
Low Speed Scanner Production 112
Pension and Bounty Land Warrant File Sizes 118
GPO File Sizes from High Speed and Low Speed Scanners 119
GPO File Sizes at Different Scanning Resolutions 119
Image Sizes of Thomas S. Steele File from Paper 122
Image Sizes of Thomas S. Steele File from Microfilm 123
Image Sizes of John Steinart File from Paper 123
Image Sizes of John Steinart File from Microfilm 124
Optical Disk Utilization 134
Comparison of Storage Requirements 134
Staff Search Time Test 137
Staff Search Accuracy Rates 138
Workflow Data from System Model 157
Image Compression and File Sizes 172
High Speed Scanner Specifications 211
Mode Menu Options 225
Laser Printer Specifications 264
Quantity of CMSR Records Converted 282
Ranges of CMSR File Sizes 283
High Speed Scanner Production 284
Indexing Production 284
Quality Control Production 285
Low Speed Scanner Production 285
Images Written to Optical Disk 286
Average Daily Production Rates - All Work Days 288
Average Daily Production Rates - Active Days 288
Generic Study Model 291
Paper System - Cost Breakdown 296
Manual Microfilm System - Cost Breakdown 300
CAR Microfilm System - Cost Breakdown 306
Digital Image/Optical Disk System - Cost Breakdown 310
CAR Microfilm System with Conversion by Service Bureau
Cost Breakdown 314
Digital Image/Optical Disk System
with Conversion by Service Bureau - Cost Breakdown 318
Cumulative Cost Comparison 320

PREFACE

The Optical Digital Image Storage System Project Report is the culmination of a five-year effort. In 1984, the Archivist of the United States approved the undertaking of a pilot project to research and test the application of digital imaging and optical disk technologies to archival programs. During the course of the project, a digital image capture and retrieval system was designed and procured, and experimentation with the processing of a broad sample of archival documents was conducted.

This report consists of six chapters and nine appendices. Chapter 1 is a management summary which places the overall contents of the report in an archival context. It also presents the goals and objectives of the project and summarizes the conclusions. The remainder of the report provides detail on the technology and its application to archival programs.

Chapter 2 traces the chronology of the project from its conception through the acquisition of the pilot system used in the test. Chapter 3 discusses current archival programs for preservation and reference service at the National Archives so that readers can understand the context in which application of automated digital image systems could occur.

Chapter 4 provides brief descriptions of the equipment acquired for the ODISS pilot system. Full detail of the system and its operation is provided in Appendix B. Chapter 5 presents the ODISS operational test plan which was the "blueprint" by which the project testing was conducted. Chapter 6 describes in detail the actual project testing and presents the technical findings.

Appendix A provides an overview of digital imaging and optical disk technologies for the reader who is new to the subject. Appendix B describes the ODISS subsystems and their operation in considerable detail. Appendix C summarizes the system quantitative and performance data compiled during ODISS testing. Appendix D presents a cost analysis which compares various document conversion options based upon a generic model.

Appendix E lists the data collection forms which were used during testing phase of the project and which are referenced in the text of Chapter 6. Appendix F presents implementation considerations in undertaking the ODISS project and setting up the research system. Appendix G describes the National Archives' current use of micrographics technology. Appendix H contains photographs of the various ODISS workstations. Appendix I contains a glossary of technical terms.

CHAPTER ONE

ARCHIVAL MANAGEMENT
AND
TECHNOLOGY SUMMARY

1 ARCHIVAL MANAGEMENT AND TECHNOLOGY SUMMARY

1.1 Project Origin

In February 1984, the National Archives issued *A Study of Alternatives for the Preservation and Reference Handling of the Pension, Bounty-Land, and Compiled Military Service Records in the National Archives*, which recommended conversion of those holdings to digital images stored on optical disks with an automated, indexed retrieval system. Later in 1984, the National Archives issued *Technology Assessment Report: Speech Pattern Recognition, Optical Character Recognition, Digital Raster Scanning*, which recommended that the National Archives evaluate the feasibility of using digital imaging and optical disk storage technologies. Six months later, the Archivist of the United States formally approved a research test called the **Optical Digital Image Storage System**, otherwise known as **ODISS**.

1.1.1 Goals

The Archival Research and Evaluation Staff undertook the ODISS project to demonstrate and evaluate the feasibility of digital imaging, optical disk storage, and computer retrieval technologies as alternative conversion, storage, and retrieval technologies for the National Archives. Among the goals of the project were the following:

- To establish the feasibility, costs, and benefits of converting paper and microform documents to optical digital media and to assess document input speeds required to accomplish conversion in an operational environment

- To determine the optimal scanning density for documents consistent with producing legible images while minimizing storage requirements

- To assess the storage capacity of the system and media in terms of storage cost and efficiency

- To evaluate system capability to automatically retrieve stored optical images using electro-mechanical devices

- To determine the suitability of creating printed document images from digital data

- To determine staff and public reaction to and acceptance of an image retrieval system as opposed to the paper and microfilm currently used for reference

1.1.2 Test Sample

A key consideration in designing the ODISS project was the selection of archival material to be used in the test. Satisfying the project goals listed above led to the following criteria.

- The archival records converted in the ODISS project should exist in both paper and microfilm in order to compare digital and micrographic conversion technologies.

- The archival records should be representative of general document types and characteristics of other bodies of records in order to generalize about the feasibility of digital imaging technology.

- The archival records should be fairly active in order to evaluate staff and public reactions to image legibility and system retrieval performance.

- The archival records should be of reasonable volume in order to complete the conversion within approximately six months.

Based upon these criteria, the Tennessee Confederate Compiled Military Service Records (CMSR) were selected. These records exist in paper and microfilm, the latter having been filmed in the late 1960's.[1] The Tennessee CMSR, which total about 400 cubic feet, are a relatively small portion of a larger body of compiled military records which total 30,000 cubic feet. A GSA survey of the CMSR in 1983 estimated that the average CMSR file contains 15 page images.

Because the Tennessee CMSR are representative only of other comparable CMSR, pension, and bounty land records, the experience with their conversion cannot be used to generalize with great confidence about the feasibility of using digital imaging technology to convert other archival holdings whose attributes might differ substantially. Using the results of a 1985 preservation holdings survey, test documents were selected that were considered representative "problem" document types with poor quality images. These tests, frequently referred to in this report as *ad hoc* tests, were intended to provide a reasonable basis for extrapolating ODISS findings to National Archives' holdings in general.

Having an optical digital image storage system test facility offered a number of advantages, including the possibility of using state-of-the-art technology at a fraction of the cost of a full production system. Nonetheless, as with any other test project, it was inevitable that the specific hardware and software environment in which the test was conducted would have limitations. Therefore, this report distinguishes between general capabilities of optical digital image technology for capture, storage, and retrieval; the specific capabilities of the ODISS hardware and software configuration; and the tools for production management discovered during the Tennessee CMSR conversion.

1.2 Technology Summary

The key technologies involved in the ODISS Project are digital imaging, optical storage, and computer retrieval of document images.

1.2.1 Digital Imaging

Digital imaging is an electronic process whereby an image of a document is captured. It involves scanning devices which measure reflected light from document pages and convert the measurements into digital information. Typically, scanners can capture images at 200, 300, and 400 pixels or dots per inch. A pixel, or picture element, is a discrete point on a document whose lightness or darkness is sensed by the scanning equipment. Image sharpness usually is improved as the number of pixels per inch is increased. (For a more detailed

[1] Filming of the Tennessee CMSR was done by a contractor in the late 1960's. Apparently, the filming was not done in accordance with existing standards at the time, which resulted in poor quality film images. Since images captured from degraded microfilm could not be expected to compare adequately with those captured from original paper documents, several controlled tests using current microfilm technology were utilized. See Chapter 5 of this report for a discussion of the test plan.

explanation of pixels and digital imaging technology in general, see the discussion in Appendix A.) Digital imaging technology also involves the use of "enhancement algorithms" to improve the legibility of images. Image enhancement is particularly useful in "cleaning up" stains or intensifying poor resolution page images to make them more legible.[2]

1.2.2 Optical Disk

Optical disk storage technology involves recording information at extremely high storage densities.[3] This recording technology employs optical lenses to focus a laser beam down to an area measuring several microns.[4] The on and off state of the laser beam represents the stream of pixel elements (ones and zeros) in a scanned page image. When the laser beam is on, it physically alters the data-sensitive area of a specially fabricated disk. When the laser beam is off, no change occurs in the data-sensitive area. In one major category of optical disk technology, the resulting physical alteration is not erasable, which makes it a much more desirable archival storage medium than magnetic media. This non-erasable feature has given rise to the term "WORM" disk, which means <u>w</u>rite <u>o</u>nce, <u>r</u>ead <u>m</u>any times.[5]

1.2.3 Computer Retrieval

The computer retrieval technology employed in the ODISS project uses index information stored on a magnetic disk to conduct searches for files that meet user-specified criteria. The index information on individuals documented in the CMSR files is linked to digital page images stored on optical disks. When a search of the index verifies that a file on an individual exists, the page images on optical disks can be retrieved for display or printing.

1.3 Archives and Management Issues

1.3.1 Background

The primary mission of the National Archives is to preserve the permanently valuable records of the federal government and to make these records available to users. Achievement of this archival mission, of course, must be carried out in a cost-effective manner. Consequently, this section identifies critical archives and management issues and within this context discusses major findings of the ODISS Project.

The permanently valuable records of the federal government now in the National Archives comprise about 1.6 million cubic feet, the vast majority (an estimated three billion pages) of which are paper-based. Over the years, the National Archives has performed large-scale

---

[2] Technical aspects of image enhancement are discussed in Appendix A, while archival considerations are noted later in Section 1.3.2.4. Image enhancement also is defined in the Glossary in Appendix I.

[3] Two-sided 12-inch optical disks are available today that can store 4.8 gigabytes of information. It would take approximately 27 reels of standard half-inch tape written at a density of 6250 bpi to store the same information.

[4] A micron is one millionth of a meter, or one thousandth of a millimeter.

[5] See Appendix A and the Glossary in Appendix I for more detailed explanations.

document conversion (usually with microform technology) in order to limit physical handling of original paper documents, to reduce storage space requirements, or to distribute records to multiple sites. Of these three objectives of the National Archives' document conversion program, the limitation of physical handling is clearly rooted in the National Archives' archival preservation mission. Therefore, any evaluation of a document conversion technology must take into account the preservation aspect of archival management.

Both the International Council on Archives and the National Archives have identified the primary archival requirements for a document conversion program, which encompasses preparation of the documents for conversion, the actual conversion process itself, the utility of the converted images, and the stability of the new storage medium.[6] Thus, for the purposes of this summary, these archival requirements for a document conservation program form a significant part of the context in which the findings of the ODISS project are presented.

As noted earlier, a primary mission of the National Archives is to make records from its holdings available to users. This implies, of course, the existence of an information retrieval capability which facilitates the identification and retrieval of records. The primary tool that archivists and researchers use is a finding aid, which typically describes records at the series level.[7] Critical archives access issues for evaluating automated finding aids are speed and relevance of retrieval, simplicity of the user system interface, enhanced retrieval capability, and decentralized distribution. These archival access issues and their implications for the ODISS project comprise the second set of the archival requirements discussed in this summary.

Few archival institutions, particularly the National Archives of the United States, have unlimited funds. Sound archival management, therefore, requires that the mission of the National Archives be carried out in a cost-effective manner The key cost-benefit issues for document conversion and access technologies are document throughput, space reduction, and improved staff access to and retrieval of relevant documents for researchers, and retirement of originals from active use and into an environmentally stable storage area. Accordingly, the third major section of this summary examines cost-benefit aspects of the ODISS project within the context of archival management concerns.

1.3.2 Document Conversion Issues

1.3.2.1 Document Preparation

As noted earlier, document preparation considerations are part of the archival context of document conversion. Document preparation at the National Archives includes flattening folded papers, removing fasteners such as staples and paper clips, and correcting misfilings

[6] Frank B. Evans, *The Selection and Preparation Of Records For Publication On Microfilm*, Staff Information Paper 19 (National Archives and Records Service, 1970); and Albert H. Leisinger, Jr., *Microphotography For Archives*, (International Council on Archives, 1968).

[7] Some finding aids provide descriptions at the item or case file level. Generally, these finding aids were created by the organization that produced the records. The finding aid for the CMSR records used in this test describes the records at the case file level.

so that documents are arranged in the correct order. Fragile documents or those with special problems require special handling.

The basic document preparation issues confronting the ODISS project were whether standard document preparation practices could be used, what special provisions were necessary for preparing fragile documents for scanning, and what impact these provisions had on the document preparation activities.

Preparation of the Tennessee Confederate CMSR records for digital scanning generally involved files on individual soldiers. Documents were removed from their envelopes or jackets and placed in new folders. The documents were rearranged so that jackets were first, followed by standard size regimental cards, and any other documents. Usually, where there were other documents in the file, they tended to be tri-folded and had to be unfolded and flattened.

The standard document preparation standards and procedures used in document conversion (microfilming or electrostatic copying) required minor modification. Because the high speed scanner could scan both sides in one pass, plastic clips were placed on two-sided documents to alert the operator. Where possible, documents after the cards were arranged by size from smaller to larger. Of greater importance, however, was the placement of fragile documents in polyester sleeves to facilitate high speed scanning. Experiments were conducted to test clear polyester folders sealed on one edge, two adjacent edges, two opposite edges, and three edges. Static electricity in polyester sleeves sealed either on two opposite sides or on three sides made it difficult and time-consuming to insert a document. This was not the case with folders sealed only on one side. Nevertheless, the use of polyester sleeves did not impede the staff in maintaining their standard production rate.

> CONCLUSION. Except for the insertion of fragile documents into polyester sleeves and rearrangement of documents by size, standard document preparation procedures can be used for digital image scanning. Neither rearrangement of documents by size or inserting documents into polyester sleeves significantly affected document preparation production rates for the Tennessee Confederate CMSR, although they might for other holdings where there are substantial numbers of fragile documents.

1.3.2.2 Image Capture

A basic requirement of archival image capture is that it should not damage or otherwise cause any deterioration to the documents being copied, particularly fragile documents. Undamaged original documents are necessary for 100 percent visual comparability of originals to copied images. In addition, it is absolutely imperative that image capture should cause no damage to documents of intrinsic value.

Although the actual process of electrostatic copying and microfilming generally causes no damage to documents, the policy of the National Archives is to avoid the use of a mechanical paper transport because of the potential for damage.[8] Fragile documents and intrinsically

[8] A mechanical paper transport is used in microfilming only in the case of heavy card stock which is not likely to be damaged.

valuable documents are individually positioned and filmed under a glass platen. Of course, special handling for these documents results in a substantial reduction in production.[9]

Neither the actual digital image scanning process nor the mechanical paper transport of the high speed scanner caused any damage to documents. The paper transport was flexible enough to allow for a variety of document types and conditions. An air vacuum held documents flat against transport belts as the belts passed under the scanner. When the vacuum was released, documents gently dropped down into a hopper. Fragile documents were placed inside polyester sleeves which were fed into the high speed scanner and scanned with no difficulty. Paper transports for electrostatic copiers typically cannot handle thick material like polyester sleeves. Although polyester sleeves could be used with a vacuum belt paper transport for a high speed microfilm camera, the high reflectance of the polyester film would cause severe image problems.[10]

> CONCLUSION. The ODISS mechanical paper transport caused no damage or deterioration to any of the Tennessee Confederate CMSR records. The high speed scanner processed with no difficulty fragile documents placed inside polyester sleeves.

A second archival consideration for image capture is that it should allow for image replacement in order to correct any problems, such as images missed in the scanning process or images of unacceptable quality. In electrostatic copying and microfilming such problems usually are identified in a quality control review that occurs some time after conversion. This is particularly true for microfilm because of the chemical processing required before images are readable. In a narrow sense, neither electrostatic copying nor microfilming provides replacement flexibility. However, in a broader sense they do, because missing documents or defective images can be recopied or refilmed. Replacement with electrostatic copying involves manually removing the inadequate image and inserting the recopied one. For microfilm, splices can be used to replace missing or inadequate images, but this is both cumbersome and inefficient.

Because digitally scanned images consist of electronic pulses (until written to an optical disk), replacement is easy and efficient. Electronic images can be erased or moved with ease, and they can be inserted into a file wherever this is necessary. Although the dynamic nature of digital image scanning permits image correction on the fly, higher scanning throughput rates can be achieved only if this is done during quality control, as it is typically done with electrostatic copying and microfilming. Once the images are written onto optical disks, the write-once nature of the media requires that any replacement images be placed on unused portions of the disks. The images being replaced cannot be physically erased from the disks, but are "logically discarded" by changing the optical media's storage index (which is kept on

[9] The production rate of microfilm camera operators for documents requiring special handling is 875 pages per day compared to 3331 images requiring no special handling.

[10] The combination of intense light of microfilming systems and the inherently high reflectivity index of polyester material accentuate the polyester film's glare and reflection characteristics. This usually results in low image contrast.

modifiable magnetic storage) so that from then on it will effect retrieval of the replacement images rather than the discarded images.

Even if a problem should occur with an optical disk, there still would be no need to rescan the documents. Data contained on the security backup disk could be copied to a new optical disk with no loss. This procedure is similar to NARA's current preservation practice of periodically recopying magnetic tapes containing electronic records.

CONCLUSION. Replacement of digitally scanned images is likely to be easier and more efficient than that of electrostatic copying or microfilming. Security backup optical disks can be used to create new and perfect copies without the need to rescan documents.

1.3.2.3 Image Utility

A third archival consideration for image capture is image utility. Image utility may be considered from two different viewpoints, one of which is the overriding concern to create an exact facsimile of a document in which no detail or physical feature such as stains, damaged areas, color, erasures, or the like is lost. From the other viewpoint, the overriding concern is to produce a legible reproduction of a document, even if this means some loss of detail or physical characteristics.

There are compelling arguments for both perspectives of image utility. For some researchers, where documents hold a significance of their own, e.g., documents containing famous signatures or those that represent an historic event such as a treaty, document conversion must yield an exact duplicate. Also, where document authenticity is critical, it may be important to distinguish between colors of ink in signatures, erasures, stains, and the like. For these researchers, any document conversion that eliminates such detail should be avoided. For other researchers, image legibility is far more important and the capacity to improve the legibility of poor quality original images by removing stains or intensifying contrast, for example, is a benefit. For these researchers, it makes little sense to reproduce a faint and virtually unreadable document when it is possible to make that same image readable and therefore useful. Choosing between these two alternatives is not easy, although conventional archival wisdom leans toward improved image legibility.[11]

All document conversion processes involve the loss of some physical features from the original. The only conventional document conversion process that does not cause any perceived loss of physical features is high density color photography, which typically is used for publication of high quality images. This expensive process is impractical as a production tool. Microfilm, of course, can produce exact gray scale images (no color), particularly when special film and processing are used. Current electrostatic copy technology can capture color and in fact produces slightly intensified images.

[11] Trudy H. Peterson, Assistant Archivist for the Office of the National Archives, believes that the vast majority of researchers are more concerned with legible facsimiles than with exact facsimiles. In her view, the appropriate way to handle this issue is to produce legible facsimiles of all documents and retain original paper documents of intrinsic value for the handful of researchers who may find it necessary to consult the originals.

Digital image scanning technology can capture documents without any loss of pertinent detail, including color and shades of gray.[12] However, color scanners currently are very expensive and the storage and display requirements for digital color and gray scale images are prohibitive for large image databases. Because of these cost considerations, the improvement of image legibility without the use of color or gray scale became the primary goal of the ODISS project.

Improvement of image legibility is a function of the scanning density and the enhancement algorithms used. Unfortunately, an acceptable rigorous methodology for determining what constitutes adequate image legibility does not exist.[13] In order to establish baselines of image legibility, a number of documents were scanned at 200, 300, and 400 dots per inch. These images were used to develop a staff consensus of acceptable image legibility. These experiments indicated that 200 dots per inch produced images of acceptable readability to the ODISS project staff. This scanning density was used both in high speed scanning and low-speed scanning.

In addition to the project staff's consensus about acceptable image legibility, a number of National Archives' reference support staff reviewed scanned test documents and were asked to rate the images from unacceptable to highly acceptable. These survey results confirmed that in most instances a scanning density of 200 dots per inch along with appropriate enhancement algorithms produces acceptably legible images.

The high speed scanner employed a constant thresholding image enhancement algorithm that could be invoked by operator selection of one of eight contrast levels. This algorithm, along with a scanning density of 200 dots per inch, proved sufficient to capture excellent Tennessee CMSR images in 94 percent of the cases. The remaining six percent of the images were scanned on the low-speed scanner with equally impressive results.

Density levels higher than 200 dpi are usually required only for instances where capture of minute detail is mandatory. This requirement is generally found in cartographic material at small scale, engineering drawings, and type sizes of less than five point scale. Most other collections could be digitized at 200 dpi which would result in completely legible images. However, scan density is not the only factor influencing image quality. Enhancement techniques can help create a clearer, more defined image without increasing scan density. This technique not only provides more detail to the image, but also accomplishes it without increasing image file sizes.

---

[12] The approach of the National Archives of Spain in its optical disk project calls for full gray scale capture of some nine million images of documents that deal with the Spanish Empire from the time of Christopher Columbus to the early 20th century. Capturing and storing images in full gray scale will enable researchers to view exact duplicates (excluding color, of course) of the documents and to determine for themselves what constitutes useful information. As noted below, there are considerable equipment, processing, and storage costs associated with capturing and storing this level of detail of scanned images. The ODISS project involved documents of lesser vintage, most of which have no intrinsic value. Consequently, cost considerations led to the decision to store "binarized" images.

[13] According to *Preservation Microfilming. A Guide for Librarians and Archivists*, (1987, Nancy E. Gwinn, ed.), "The creation of legible images on microfilm remains an art rather than a pure science. Various standards recommend ways to produce sharp, legible images, but they are not totally scientific and depend in part on the camera operator's judgment." (p. 102)

The low-speed scanner image enhancement algorithms[14] delivered by the original contractor proved somewhat ineffective in producing legible images from seriously degraded documents. Fortunately, an IPT Scan Optimizer,[15] with patented algorithms specifically designed to improve stained, faded, and low contrast documents, was loaned to the National Archives on a beta-test basis and installed in the low-speed scanner. Images enhanced at the low speed scanner by the Optimizer algorithms were very readable, even when the originals were barely legible.[16] A major benefit of using the IPT Optimizer was that its images contained less "noise," or irrelevant background information and therefore required less storage than those produced by the unimproved low speed scanner.

Another image enhancement technique resulted from experimentation with the color of the background reflectance surface during scanning. The low speed scanner, as delivered, used a white reflectance surface which worked well with most CMSR documents, but not with two-sided documents having significant ink bleed-through. It was found that red, blue, and brown reflectance surfaces greatly improved the legibility of documents with bleed-through problems.[17]

Non-CMSR documents were used to evaluate problem document types scattered through the holdings of the National Archives. A 1985 National Archives' preservation holdings survey identified document types that included preprinted material, different colored sheets, tissue paper, dark and light inks, colored ink stamps, blurred carbon type, faint pencil handwritten notations, turquoise carbon ink on translucent paper, purple carbon on brownish paper, and blue carbon on buff colored paper. Selected documents identified in the survey were scanned at 200, 300, and 400 dots per inch on the low speed scanner with and without the IPT Optimizer. Generally, a scan density of 400 dots per inch produced very sharp images, especially fine line detail and character edges. However, these same documents, when scanned at 200 dots per inch, yielded legible images and with significantly lower storage requirements, typically as much as one-third that of images scanned at 400 dots per inch.

An eight-bit gray scale scanner was included in the ODISS equipment configuration. Although the gray scale scanner is a very powerful tool, it was not very practical for routine ODISS production operations. In general, gray scale scanning increased storage requirements by up to 800 percent. Using this software for image enhancement, the gray scale scanner took between five and ten minutes to scan one full page image. Transfer of this bit stream to the low speed scanner for entry into the CMSR system required several minutes. This combination of high storage requirements and long processing time is impractical for general use. Of course, where it is essential to produce exact duplications of original documents, a gray scale scanner is necessary. In the ODISS environment, however,

---

[14] Thresholding and textual removal.

[15] Processor board and accompanying control panel developed by Image Processing Technologies of McLean, Virginia.

[16] Figure 6-3 and Figure 6-4 in Chapter 6 illustrate the difference in image legibility produced by the IPT Optimizer.

[17] This parallels the experience of the Spanish National Archives optical disk project where black and brown reflectance surfaces greatly improved image legibility.

the chief benefit of the gray scale scanner was in evaluating enhancement algorithms and demonstrating enhancement technology to visitors.

Because the ODISS research test focused upon black and white text images, it was not possible to evaluate the image utility of color-based documents or color photographs. This is an important area of digital imaging technology that requires further research and investigation.

Another important aspect of image utility involved scanning microforms. The ODISS equipment included an experimental multiformat microform scanner, which was used to evaluate image utility of documents scanned from both poor image quality and high image quality microfilm. As a baseline of comparison, identical images from original paper files and microfilm copies[18] were scanned at 200, 300, and 400 dots per inch on the low speed scanner and the microfilm scanner. The resulting scanned images were displayed on a high resolution terminal screen and printed on a laser printer. Images from the original paper documents scanned on the low speed scanner were more legible than those from the microform scanner. One reason for this is that the latter did not utilize the IPT Optimizer.

A major factor contributing to acceptable legibility was the use of enhancement algorithms. The available thresholding algorithm did improve legibility. Image enhancement of extremely poor quality microfilm images, however, did not produce miracles. Even after image enhancement, these digital images were of marginal quality at best. Of course, it is quite likely that installation of the IPT Optimizer in the microform scanner would have led to drastic image improvement. Further investigation and research in this area could confirm if significant image improvement of extremely poor quality microfilm images is indeed possible.

The key issue, however, for digital image conversion of poor quality microfilm images in the National Archives is that much of the older film lacks blip marks, which are essential in high speed scanning. Without the capability of automatic frame alignment, each microfilm image would have to be hand positioned. There are some prospects that a mechanical film transport under computer software control could be developed to permit high speed scanning.[19] Given the availability of a high feed mechanical film transport, it is clear that digital image scanning of poor image quality microfilm as opposed to the original records (where they exist) would produce legible images.

---

[18] Because the microfilm copy of the Tennessee Confederate CMSR is of very poor quality, it was used to verify if adequate image utility could be achieved in such adverse circumstances. If adequate image utility could be achieved, then a large-scale production conversion of the CMSR and other microfilm holdings would be feasible. Use of microfilm would eliminate the need for document preparation and, with a mechanized film transport, would greatly accelerate scanning throughput.

[19] Some exploratory work has been done on software that can detect (i.e., find) images on microfilm that has no frame registration method (i.e., no blip marks or sprocket holes). The Archival Research and Evaluation Staff has had exploratory discussions with several vendors who are interested in developing a prototype scanner. In addition, the staff completed a requirements analysis for such a capability in March, 1988. See *Sanitized Documents Image Reference System: Requirements Analysis* by Michael Goldman.

CONCLUSION. Digital imaging technology can deliver legible images and even significantly improved images when required. A scanning density of 200 dots per inch was adequate for producing legible images from more than 98 percent of the CMSR records scanned, particularly when linked to enhancement algorithms of the IPT Optimizer. Although digital imaging technology cannot produce legible images from illegible microfilm images, nonetheless it can produce legible images from both good and extremely poor quality microforms. Additional research and development in automated film transports must be completed, however, before production-level digital scanning of unblipped microfilm should be considered.

1.3.2.4 Image Stability

As noted earlier, a primary purpose of a document conversion project is to extend the life of information of archival value, which for the purposes of this report means image stability over time. Image stability is a function of the longevity of the storage medium and retention of all information captured in the conversion process. The latter is particularly important when copying images onto similar or other storage media from one generation to another. An evaluation of the image stability features of optical disk media must take into account comparable features of paper and microfilm.

The longevity of high quality paper and silver halide microfilm is impressive. Documents written on paper with a high cotton content or copied on silver halide microfilm are quite likely to have a useful life of "at least several centuries" when stored in the proper environment.[20] Documents copied on diazo microfiche or low quality paper such as Therma-Fax are not likely to last more than ten years, if exposed to heat or light. Of course, newspaper has a relatively short life also. A critical factor in the longevity of both paper and microfilm is usage. Even the highest quality paper and silver halide microfilm will deteriorate if subjected to high use. This is particularly true for microfilm which is vulnerable to scratches resulting from continued use in microfilm readers. However, image stability of microfilm can be assured by retaining a master negative copy of the microfilm.[21]

The Sony Corporation, the supplier of the optical disk media used in the ODISS project, claims longevity for their media in excess of one hundred years. Of course, it was not possible for the ODISS project to verify this claim. However, in a parallel project, the National Archives helped to establish a National Institute of Standards and Technology (NIST) Optical Media Laboratory to develop a standard testing methodology to predict the life expectancy of optical media. The initial NIST testing was done using Sony optical disks and a progress report was completed in September 1989. The progress report included preliminary data for 60°F and 80°F aging environments. If confirmed by a third point, this

---

[20] Adelstein, P.Z. and McCrea, J. L., *Stability of Processed Polyester Base Photographic Films*, Journal of Applied Photographic Engineering; (December 1981); pg. 160-167.

[21] The National Archives maintains a master silver halide negative as well as a master reference silver halide negative copy. Only the master reference microfilm copy is used to produce working copies of microfilm.

would indicate a life expectancy from 30 to 130 years.[22] The longer life expectancy reflects test conditions where the relatively few error bursts of great length are excluded from the calculations.[23]

Unlike paper and microfilm, repeated use of optical disks causes no loss of information. The principal reason for this is that in the process of "reading" an optical disk a low power laser beam is used and there is no physical contact between the data sensitive area on the disk and the reading mechanism. As noted earlier, this capacity of optical disks to be read many times without loss of information has given rise to the term, "WORM disk." During the ODISS project, several CMSR files stored on the Sony optical disks were read several thousand times with no loss of information.

In this regard, another attractive feature of optical disks is that they do not require special environmental storage conditions. Being exposed to ambient temperatures poses no long-term problem for these disks.

A critical part of image stability over time is the potential for information loss in generational copying.[24] For example, both electrostatic and microfilm generational copying cause loss of physical detail, typically involving degraded contrast or resolution. In both electrostatic and microfilm copying, there is some loss of resolution or contrast.[25] This means that in the third generational copy, a substantial part of the resolution of the original document has been lost. Of course, significant generational loss of image information generally is not a problem as long as the first generation copy is recopied, as in the case of making microfilm copies from master negatives of high quality.[26]

Optical disk media have been subjected to repeated generational copying with no loss of information. In the ODISS project, copies of several CMSR files stored as digital images on the Sony optical disks were retrieved and then written again on the disks. This process was repeated through ten generational copies with no perceived degradation of the images or loss

---

[22] The NIST report, *Development of a Testing Methodology to Predict Life Expectancy of Optical Disk Media*, (1989), cited the need for additional data, including a third point for verifying this use of the Arrehnius theory. The progress report also stipulated that the "Arrehnius plots showing life expectancy values are included with the sole effect to illustrate the testing approach" and that the program was not "aimed toward evaluating a particular type of media. Commercially available media have been used for the tests with the sole intent to evaluate the test procedures implemented." A published version of the NIST report with additional test data is expected by the spring of 1991.

[23] *ibid*, p. vii.

[24] Generational copying means making a copy of an original, and thereafter, making a copy of the copy. The process may be repeated indefinitely with the key point being that each new copy is always made from the most recent generational copy.

[25] Albert H. Leisinger notes that "when a film copy is made from the camera negative, there is always a copying loss, sometimes as much as 20 percent." *Microphotography For Archives*, p. 17.

[26] The CMSR microfilm, produced by a contractor, yielded very poor quality master silver halide microfilm, with major image problems. This is not the case with microfilming of the Navy Widow Pension Files. which was done by NARA staff.

of information between the first copy and the last.[27] In a related study, the Public Records Office of the United Kingdom copied magnetic tapes to optical disks and back to computer tapes with no loss in information.[28]

A critical concern in using optical disk media for archival storage is that of ensuring future retrievability. Unlike paper and microfilm, digital images stored on optical disks are not human-readable. Complex computer systems are required to read an optical disk and interpret the binary ones and zeros of a digital image and then display the image in a way intelligible to a human. This machine dependence introduces the factor of technology obsolescence, as in any computer-based system. Use of optical disks in archives storage will require the capability over time to transfer digital images from one system to another. Recopying of optical disks from one system to a newer version of that system before the older version is no longer supported by its manufacturer will prevent loss through technology obsolescence. However, the capability to transfer easily digital images written on optical disks using a data format peculiar to one computer system to one that does not use the same format does not now exist. Major work is underway, however, to develop standards that can ensure the transferability of digital images between dissimilar computer systems.[29]

CONCLUSION. The stability of digital images stored on Sony optical disks is potentially adequate for one hundred years of archival storage, a period of time substantially less than that of high quality paper and silver halide microfilm. However, unlike high quality paper and silver halide microfilm, repeated use of digital images stored on Sony optical disks causes no image degradation or loss of information. The major preservation disadvantage to using optical digital disks for archival storage is the necessity to recopy them periodically in order to avoid technology obsolescence. It is expected that the development and implementation of data transfer and physical media standards for digital images and optical disk systems will lengthen the period of time between recopying.

1.3.3 Document Retrieval Issues

1.3.3.1 Document Access

Making records available to researchers is fundamental to the mission of the National Archives. To meet this objective, the National Archives uses a manual information storage and retrieval system that utilizes series descriptions as the primary tool for identification and

[27] Raw byte error rates were below one erroneous byte per 10,000 bytes. Built-in error correction codes reduced this to $10E^{-12}$ which indicates only one erroneous byte per gigabyte of data.

[28] The tapes contained numerical data, not text or image data. See the report entitled, *Public Record Office. Optical Disk Project Final Report*, 1989.

[29] The Association of Information and Image Management (AIIM) is sponsoring developmental work in this area. In addition, a special interest group from the Digital Image Applications Group (sponsored by the National Archives) also is working on the development of such standards. It will be several years before these standards are in place.

retrieval of records. In some instances, records are indexed either at the folder (file) or document level. As a general rule, ease of access to records in the National Archives is a function of whether or not individual files and documents are indexed.[30] Where the records are indexed, a specific folder can be retrieved and specific documents located. Where access is at the series level, boxes of records must be searched until the specific documents are located. Because the CMSR files are indexed, it is possible to search a name index for a specific file and then retrieve specific documents. This search and retrieval activity involves staff access and public access. Typically, the latter occurs in the self-service Microfilm Reading Room.

Neither the National Archives nor any professional archival organization has identified appropriate archival criteria for evaluating automated retrieval. Nonetheless, for the purposes of the ODISS project and this report, the key criteria are the speed and relevance of retrieval, simplicity of the user system interface, enhanced retrieval capability, and decentralized distribution. Speed, of course, refers to timeliness while relevance denotes that the retrieved information matches the user's query. Simplicity of user system interface means that no special skill or understanding is required to use the system. To phrase this another way, it means a user-friendly search and retrieval system. Enhanced retrieval capability means that the retrieval system supports forms of inquiry not available in the original indexing scheme. Finally, decentralized distribution refers to the capacity to extend search and retrieval functionalities to a wide variety of locations.

1.3.3.1.1 Speed and Relevance

Staff use of the CMSR files typically involves processing mail-in requests from researchers for information about an individual soldier and manually searching the appropriate index. If the index search identifies a soldier meeting the specified criteria, that is, the search produces relevant information, the file is retrieved (either on microfilm or paper), and a copy is made and sent to the requester. Current NARA procedures call for staff to complete an index search and retrieve the appropriate file in 9.6 minutes or less.[31] It is difficult to establish the level of accuracy of CMSR index searches because as a general rule, staff members use the exact name spelling provided. If the name provided is incorrectly spelled or there is a variant spelling (e.g., Stephens or Stevens), no file is located, even though in fact a file may exist under the correct or variant spelling.[32]

Four staff members who regularly perform index searches for Civil War soldiers participated in an evaluation of the speed and relevance of the automated index of the Tennessee Confederate CMSR. The ODISS project staff prepared six batches of ten queries each, consisting of a mixture of easy and difficult searches, to simulate the batches of mail-in queries that the reference staff answer. After a training session of two hours followed by practice time of two hours, the four staff used the ODISS automated index to respond to the

---

[30] Even if no index was explicitly developed for them, records filed in some particular order (e.g., alphabetical or chronological) are implicitly indexed.

[31] This does not include any time for reviewing or duplicating the contents of a file.

[32] If an index card notes there is a second file under a variant spelling, then this index information is also consulted.

six batches of queries. The average time to complete five batches[33] ranged from a low of 1.7 minutes per query to a high of 4.5 minutes per query, which was a substantial improvement over the manual index search and file retrieval times.[34]

The relevance or accuracy of the searches was not as dramatic. Accuracy was calculated on the basis of the number of searches that correctly identified the soldier, which ranged from a low of 50 percent to a high of 90 percent. This low accuracy rate does not reflect a shortcoming in the retrieval technology, but rather the failure of the four staff members to use cross reference index information to a second file which contained more records. For example, a search for information about Andrew Crowson, a cavalry soldier, yielded one jacket image with a remark that other cards were filed with A. J. Crowson in #31221. However, the CMSR staff member overlooked this reference and did not retrieve the second file, resulting in an inaccurate search result. Subsequent interviews with the four participants strongly suggested that it was not possible for them to learn thoroughly in four hours the flexible rules for performing index searches and document retrieval on ODISS.

The original ODISS Test Plan called for an evaluation of the experience of non-staff researchers conducting their own CMSR searches. An ODISS public workstation was installed in the Microfilm Reading Room, which is essentially self-service, to collect data on researchers' interest in and ability to learn and use a self-service automated reference station. This capability was never tested because the on-screen instructions for the public did not elaborate enough and did not provide a novice with sufficient guidance to understand easily the system's operation or navigate through its menu paths. This made it difficult and time-consuming for researchers to work through a self-teaching session.

An alternative approach was designed to collect researcher reactions to the public workstation. Three volunteers, each of whom had some previous computer experience, were assisted by an ODISS project staff member in working through the screen instructions. The volunteers rated various features of the workstation. Its most desirable features were image legibility and image manipulation, especially the zoom (or image magnification) capability. Despite the volunteers' enthusiastic response to the public workstation, it was not possible to evaluate the timeliness and relevance of their searches. Consequently, further investigation of researcher reaction to self-paced instructions for index searches and image retrieval should be undertaken before an operational search and retrieval system for CMSR records is designed.

In addition, ODISS project staff conducted demonstrations of the public reference terminal in the Microfilm Reading Room using the scanned Tennessee CMSR files. Index searches, file retrievals, and laser-printed hardcopies of images were demonstrated. Reactions to the demonstrations were quite favorable and often enthusiastic.

---

[33] One of the six batches was discarded since one staff member forgot to record his time for that batch.

[34] It should be noted that unlike the CMSR manual search, the elapsed ODISS search and retrieval times also included the display of document images which could then be printed out in a minute or so, depending upon how many images were in a file.

CONCLUSION. There was a substantial improvement in the timeliness of staff index searches and retrieval of the ODISS CMSR files over manual methods, but the relevance or accuracy of the results of those searches was less impressive. Doubtless, more training and hands-on experience of the four CMSR staff would have greatly improved their accuracy rates. The design of the screen instructions for public researchers precluded collecting data on the timeliness and accuracy of public searches of the ODISS index. These aspects of the document access component of the ODISS Project require further investigation before designing a full-scale automated search and retrieval system for CMSR records. Despite these problems, many researchers in the Microfilm Reading Room who viewed ODISS demonstrations were impressed with the system's capabilities.

1.3.3.1.2 Simplicity of User System Interface

The current manual alphabetical name searches and retrieval of CMSR files either on microfilm or paper do not require complicated procedures or a lengthy sequence of steps to follow. Although the full features of the ODISS automated search capability are not intuitively easy to use, the four CMSR staff members rated the workstation as easy to use and very fast. Despite a short training time (four hours) on the ODISS system for the four CMSR staff, they were able to operate the workstation and pick up most of the basic procedures. Undoubtedly, this limited training explains why the participants achieved only an average accuracy rate of approximately 70 percent. Interviews with the four CMSR participants strongly indicated that a few days of experience using the ODISS search and retrieval capability would have made them 100 percent proficient.

The Office of the National Archives' workstation operators assigned to the ODISS Project who worked with the ODISS system daily had no difficulty in using the search and retrieval capability. Although ease of use likely is a function of familiarity with the ODISS system and the CMSR records themselves, the workstation operators had no prior experience with either one. Yet, within a very short period of time they were fully proficient in using the search and retrieval function.

The screen designs for ODISS workstations and the public workstation were similar, but the instructions were not. Although the ODISS project staff could easily follow the menu instructions, this was not true for users of the public workstation. The experience of three volunteer researchers using the ODISS public workstation confirmed this and underscores the importance of a simple and easy-to-use search and retrieval system. Identifying the appropriate screens and menu instructions for an automated search and retrieval system designed for public usage requires a very careful and systematic investigation with significant involvement of public researchers. It would be imprudent to design an automated search and retrieval system for CMSR files without first conducting this investigation.

CONCLUSION. Both the ODISS project staff and the National Archives' workstation operators assigned to the project were able to use the ODISS search and retrieval capability with little difficulty. This was less true for CMSR staff, although they undoubtedly would have been more proficient had they received more training or spent several days using the system. Nevertheless, the screen design and menu instructions comprised an adequate user system interface for them. This was not the case for non-staff researchers using the public workstation, who found the screen design and menu instructions intimidating and difficult to use. It would be imprudent to design a production system without additional study and experimentation to identify the requirements for a useful and usable system interface for non-staff researchers.

1.3.3.1.3 Enhanced Retrieval Capability

Because manual indexes are generally sorted on a single term, it is not possible to conduct concurrent multiple term searches combining name, date, or other index information. A major strength, of course, of a carefully constructed automated index is the flexibility to combine index search terms that yield more precise results. For example, searching an index for a soldier named John Smith is likely to identify a number of soldiers with that name. However, if rank and unit are combined with Smith, then the search may be narrowed to only one individual.

The ODISS automated search and retrieval system offered this kind of flexibility, a feature that both the four staff and three public researchers found very beneficial. In addition, this automated name index gives rise to new forms of inquiry not heretofore possible. For example, searches can be conducted combining rank with length and kind of service in order to identify significant trends. A fully automated index to all of the CMSR would provide an enormously rich demographic database that could support a wide range of historical inquiry not previously possible.[35]

CONCLUSION. The ODISS search and retrieval system provides a greatly enhanced access capability to the Tennessee Confederate CMSR. Extension of this search and retrieval capability to other CMSR, Pension, and Bounty Land records is likely to lead to greater use of the records as new questions of historical research are formulated.

1.3.3.1.4 Decentralized Distribution

The capacity to extend distribution to a wide variety of locations is an important aspect of document access in the 1990's. Microfilm and microfiche, of course, are well-suited to this kind of distribution. For example, a microfilm copy of the Tennessee Confederate CMSR is

[35] See for example, Constance B. Schultz, *Daughters of Liberty: The History of Women in the Revolutionary War Pension Records*, Prologue, (Fall, 1984), pp. 139-153. In particular, Schultz argues that a computer database or index to pension files could facilitate research in women's history that is not possible now.

at the Tennessee State Archives where only an inexpensive microfilm reader is required to use the microfilm. Another factor favoring microfilm/fiche in decentralized distribution is that a copy of all of the Tennessee Confederate CMSR on 359 rolls of microfilm can be purchased for $7,180.

Although the ODISS Project did not address the question of decentralized distribution of the ODISS Tennessee Confederate CMSR, there is no technical impediment to it. It could have been done either by duplicating the 12-inch WORM disks or by providing a telecommunication link between the ODISS index and image database and researchers at the Tennessee State Archives. However, even with today's technology, decentralized distribution of the ODISS Tennessee Confederate CMSR is not practical. Duplicating the 12-inch disks would cost very little, but installation of the hardware and software at the Tennessee State Archives would be very expensive. A telecommunications link between an image terminal at the Tennessee State Archives and the ODISS index and image database is neither practical nor cost-effective. The narrow telecommunications bandwidth of common voice-grade phone lines would require almost a minute per image at a 9600 Baud rate.[36] A more practical decentralized distribution that reduces the high equipment cost and eliminates the slow image transmission rate would involve copying the automated Tennessee Confederate CMSR index to CD-ROM[37] for use with the microfilm copies of the records at the State Archives.

CONCLUSION. The ODISS Tennessee Confederate CMSR can be used in a decentralized distribution environment in much the same way as microfilm is used. However, despite major advances in optical disk technology and telecommunications, the costs of a decentralized distributed ODISS Tennessee Confederate CMSR still are prohibitive. A more cost-effective approach using a combination of microfilm and CD-ROM technology could be used for a decentralized distribution of the ODISS Tennessee Confederate CMSR.

1.4 Cost Effectiveness

1.4.1 Document Throughput

A major goal of the ODISS Project was to establish a high speed digital image document conversion throughput rate that could be compared with other conversion approaches. The ODISS functional requirements stipulated a high speed scanner capable of scanning 40 images per minute, a rate approximately five times faster than that achieved by high speed microfilm camera operators in the National Archives.[38] Although the high speed scanner was capable of processing this many images per minute, other factors came into play that significantly reduced document conversion throughput to an average of 1158 images per day

[36] 9600 Baud is the fastest transmission speed possible over ordinary voice-grade telephone lines without data compression using modem units commonly available.

[37] See Section A.2.3.4 on page 195 for a detailed discussion of CD-ROM technology.

[38] This is based upon the Special Media Preservation Branch daily production standard of 3331 images of prepared flat work.

(2.8 images per minute in a seven-hour shift), about seven percent of its rated capacity and about one-third the production rate for high speed microfilm camera operators in the National Archives with comparable archival documents.[39] These factors include a significant wait time between opening and closing files, fewer images per file than projected (requiring more file openings and closings), and operator selection of certain buttons controlling document size and contrast level.[40]

An operations research analyst and statistician from the Navy Regional Data Automation Center, under contract with the National Archives, analyzed the problem of slow document throughput. He developed a computer simulation model using timing data that excluded the wait times. Because the simulation model assumes the existence of sufficient processing capacity to eliminate all wait times at the workstations, it is possible to predict realistic throughput rates. In this instance, the predicted high speed scanner throughput rate was 3888 images a day, almost 500 more than the production standard for planetary microfilm camera operators in the National Archives.

Another factor contributing to the slow document throughput conversion rate was the need for high speed scanner operators to push one or more buttons to set scanning parameters whenever certain document characteristics such as document size or contrast changed. Elimination of these activities could also contribute to higher document conversion throughput rates. Recently, a manufacturer of digital scanners reported developmental work on a high speed scanner featuring software-driven sensing devices that automatically make these decisions, thereby significantly increasing the image throughput rate. Before a full-scale CMSR conversion project is implemented, additional research and testing of high speed scanners with software that automatically sets the contrast level, image size, and the like must be undertaken in order to develop realistic production rates.

CONCLUSION. The average daily ODISS document conversion throughput rate of 1158 pages would be unacceptable in a large-scale production. Several factors, including the smaller average file size than planned for during the system design, which caused many more file openings and closings than anticipated, and a system design architecture which couldn't compensate for it, contributed to this unexpectedly low throughput rate. However, through the use of a computer simulation model, it was possible to obtain a reliable predicted throughput rate in excess of 3800 images per day. Moreover, one manufacturer of digital scanners has reported developmental work on a new high speed scanner that incorporates software that automatically selects document size and contrast levels for each image without operator intervention. Further research and investigation of this and other new scanners in an operational environment is crucial before beginning a full production document conversion operation which would use NARA staff and NARA-purchased equipment.

---

[39] It should be noted that in one controlled high speed scanning experiment, an Army Technical Manual was processed at a throughput rate of 36 images per minute with no rescans required. The uniform size and high quality of the page images largely account for these results.

[40] For a more detailed discussion of these issues, see section 6.2.2.2 on page 92 and section 6.13.3 on page 154.

1.4.2 Space Reduction

In principle, the conversion of paper records to high density storage media reduces the amount of storage space required because the originals could then be destroyed. In practice, at least in the National Archives, no paper records that have been converted to microform have been destroyed or placed in a low-cost, off-site storage facilities. In this sense, there is no absolute space reduction resulting from a document conversion program.

Digital image and optical disk systems, as demonstrated by ODISS, offer the potential for tremendous savings in storage space requirements. The 220,000 page images scanned during the ODISS conversion represent approximately 80 cubic feet of document storage space. Sony's claims that their 12-inch, two-sided optical disks would hold the equivalent of 20,000[41] page images on each side were fully validated. These 220,000 images were stored on five 12-inch disks, which together require less than one cubic foot of storage space. However, this is somewhat misleading because the jukebox, which holds the disks, and other computer equipment occupy more than 80 cubic feet of space. Obviously, the benefits of space reduction improve as the volume of paper records to be converted increases. For example, storing 500 million page images using double-density Sony optical disks would require only 250 cubic feet for disk storage. In comparison, this volume of paper would require over 200,000 cubic feet of shelf storage space. If microfilmed, the 500 million images would require almost 7,000 cubic feet of 35mm microfilm storage.[42]

> CONCLUSION. One double-density 12-inch Sony optical disk can store around 80,000 page images, which represents a significant space reduction potential, provided the original records can be removed or transferred from the premises. However, the space occupied by optical disk drives and supporting computer equipment must be taken into account in calculating net space reduction. A net space reduction is realized when the volume of paper records converted to optical disk reaches about 1,000 cubic feet. The greater the volume of paper records converted, the greater is the net space reduction ratio.

1.4.3 Improved Access

From an archival management point of view, improved staff access to CMSR files should result in decreased search time and increased accuracy of retrieval. In the ODISS project's search and retrieval test, the average time was 2.82 minutes, more than three times faster than manual search and retrieval. This suggests that with an automated search and retrieval system in place, only five staff would be required to handle all of the 100,000 to 110,000 annual write-in requests.[43] Unfortunately, the accuracy of search and retrieval of Tennessee Confederate CMSR records vis-à-vis that of manual search and retrieval cannot be compared because reliable statistics on the latter do not exist. Nevertheless, the ODISS

[41] More recently, Sony has announced a new, higher density disk that stores the equivalent of 40,000 page images per side.

[42] Refer to Table 6-22 on page 134 for a table of comparisons.

[43] Current staffing for the manual search, retrieval, and duplication activities totals 18.

automated name index search capability is a powerful tool that experienced staff could use in increasing the accuracy and consistency of their work. Indeed, the automated name index search capability permits very complex searches not possible with the manual index.

The ODISS project could not demonstrate an improvement in access to the Tennessee Confederate CMSR files by on-site public researchers, largely because the complexity of the screen design and instruction menus made self-instruction impractical. However, the automated search capability along with screen display and immediate laser printing of retrieved images can yield a major improvement in image legibility and reduced retrieval time. The latter is particularly important because it currently takes between one and two hours to deliver a requested paper file to the Research Room.

CONCLUSION. The ODISS Project demonstrated a three-fold reduction in the amount of staff time required to search and retrieve name index files of the Tennessee Confederate CMSR. This finding can be generalized to other similar name index searches in the CMSR, Pension, and Bounty Land files. Because comparable statistics are not available for manual name index searches, it is not possible to establish an improved accuracy rate for automated search and retrieval of name index files. The complexity in the ODISS screen design and instruction menus precluded an evaluation of improved public access. Nonetheless, considerable evidence suggests that a public automated search and retrieval system would drastically reduce search and retrieval time and improve accuracy.

1.4.4 Cost-Benefit Concerns

Appendix D of this report compares the costs of an optical digital image storage system and four other alternatives against a baseline manual reference system in which original paper records are used for retrieval. For consistency, the five alternatives to the paper system each require a retrospective conversion of the paper records to an alternate medium (i.e., microfilm or optical disk). In order to provide a reasonable cost comparison, a generic model application was defined and then applied to the baseline and each of the selected alternatives:

- Continued use of an existing paper storage and retrieval system (the baseline)
- Microform conversion using existing filming and retrieval facilities
- Microform conversion using upgraded equipment and a computer assisted retrieval (CAR) system
- Upgraded (CAR) microform system using a service bureau for conversion
- Conversion by digital image capture with an optical disk system used for storage and retrieval
- Digital image/optical disk system using a service bureau for conversion

The cumulative discounted costs[44] of these alternatives over a ten-year period are listed below, ranked from the least expensive to the most expensive.

✡ Existing paper	$769,875
✡ Manual microform using existing facilities	$1,186,970
✡ Upgraded microform (CAR)	$1,550,983
✡ Digital image	$1,942,376
✡ New microform with service bureau	$2,804,459
✡ Digital image with service bureau	$3,096,712

If discounted costs are plotted for ten years,[45] none of the alternative storage and retrieval technologies which involve a conversion of the records appears to be cost-competitive with the continued operation of an existing paper-based storage and reference system. Nor does the plot of the model disclose any discernable trends that would suggest any different conclusion for the decade to follow. Yet as Appendix D points out, the cost model which is the basis for this statement only applies to the particular variables, assumptions, and constraints which were used in its construction. It is conceivable that with a different set of parameters, the results would be different. It is also conceivable that rapid advances in evolutionary technologies such as digital imaging and optical storage may result in price reductions that could totally alter the picture at some point in the near to not-too-distant future.

Because any conversion of records from paper to an alternate medium involves an expensive, labor-intensive operation involving document preparation, handling of the records during the filming or scanning, and creation of an index, conversions will always suffer a cost disadvantage in comparison to maintaining and using existing holdings of records. Secondly, performing a conversion usually entails significant capital costs in the early years when equipment and services must be acquired to effect the conversion and support reference operations using the new format.

Even if the conversion of paper holdings with a labor-intensive, manual reference system to a technologically improved system involving an alternate storage medium and automated reference procedures cannot be justified currently on the basis of cost alone, there may be other intangible benefits that would warrant undertaking the conversion. Many of these benefits have been identified earlier in this chapter.

One of the primary benefits of digital imaging technology that became so obvious during the ODISS pilot is the opportunity to electronically enhance the legibility of images captured from poor-quality documents. The results in this area were quite dramatic and drew considerable favorable comment from the over 2000 persons that toured the facility during

[44] For a discussion of the use of *cumulative discounted costs* and *net present value*, refer to section D.1.2 and footnote #106 on page 292, and section D.8 page 319.

[45] Refer to Figure D-1 on page 321.

its operation. Furthermore, digital images are in fact digital data and may be copied from generation to generation with no loss of data or detail.

Whereas current reference procedures for the CMSR documents involve time-consuming and labor-intensive manual searches of indexes and subsequent manual retrieval of the records for review, ODISS demonstrated the capability for researchers or archives staff to perform the entire reference operation from a single workstation where the user would search an automated, computer-based index, retrieve and view the selected file images, and ask for paper replications as desired. With the CMSR files available on-line and available for concurrent access from more than one workstation, there was never any "out-of-file" or "lost file" situation. Staff and public researchers who used the system or were showed demonstrations were seemingly impressed and pleased at the prospect of rapid, accurate access to the files. Hence, in some cases, a conversion to a fully automated format may be warranted for heavily referenced collections in order to provide a better level of service to the public and to reduce reference staff workload.

Conversion of large holdings of paper documents to a more compact form such as microform or digital images on optical disk will result in a savings of storage space at the reference facility as long as the original documents are retired to off-site storage. If the lack of storage space at the reference location is of paramount concern, then a conversion may be warranted.

All documents suffer from repeated handling. In the case of older, deteriorating, or degraded documents, conversion may be warranted as a preservation measure in order to remove the original documents from active reference usage and retard or prevent their further degradation.

CONCLUSION: From the cost analysis presented in Appendix D, it appears that any conversion of paper records to an alternate form cannot currently be justified purely on the basis of cost alone. However, the ODISS project also identified other intangible benefits, such as improved image legibility, improved timeliness and accuracy of access, an enhanced retrieval capability, reduction of storage space requirements, and reduced or eliminated handling of original documents. A conversion of records to an alternate form may be justifiable on a basis other than reduction of costs. Each case must be decided on its own individual merits.

CHAPTER TWO

PROJECT HISTORY AND PURPOSE

2 PROJECT HISTORY AND PURPOSE

This chapter discusses the ODISS project from its conceptual beginnings in 1984. It also presents the original project goals and objectives, and details the acquisition and system implementation processes.

2.1 Origins of the ODISS Project

ODISS stemmed from a February, 1984, report from the Archival Research and Evaluation Staff (NSZ), *A Study of Alternatives for the Preservation and Reference Handling of the Pension, Bounty - Land, and Compiled Military Service Records in the National Archives.* The report evaluated technological alternatives for preserving and performing reference service on 80,000 cubic feet of military service, pension and bounty land records for which there were more than 100,000 mail-in reference requests annually. The study recommended records conversion with an automated index and raster scanned images stored on optical disks. A silver halide microfilm copy from the raster scan would be stored off-site as a security copy.

The report anticipated a full scale conversion to optical disks within seven years. However, NARA decided on a more cautious approach, because many questions remained unresolved about the feasibility of obtaining high quality images from archival documents with faded writing, various combinations of colors of paper stock and inks, and other legibility problems. The speed of conversion, the ease of training operators to run a conversion system, and the researchers' acceptance of digital images on computer terminals with automated indexes were other unanswered questions.

This cautious approach is reflected in the NSZ October, 1984, *Technology Assessment Report: Speech Pattern Recognition, Optical Character Recognition, Digital Raster Scanning.* This report described each of the three technologies and then assessed their applicability to archival records. It concluded that all three should continue to be tracked, but that OCR and digital raster scanning were sufficiently developed that NARA could usefully undertake research projects testing these two technologies with archival records.

Immediate full scale digital conversion of the pension, bounty land and military service records was abandoned in favor of a research test project. On March 6, 1985, the Archivist of the United States formally approved a research project in which NARA would acquire a test system. This system would allow research and experimentation into the many questions about the feasibility of applying digital image and optical disk technologies to National Archives holdings.

2.2 Project Objectives and Procedures

The ODISS research test was designed to take full advantage of the latest commercial advances in the electronic digital imaging technology as applied to source document and microform capture, indexing, enhancement, storage, display, and hardcopy output. ODISS enabled NARA to investigate the ability of new technologies to solve current records management problems, as well as aid in decisions regarding larger, more complex systems in the future. Some of the objectives and procedures pursued during ODISS testing included the following.

Preservation: Refine the estimates of the repair and conservation workload needed to prepare documents to the image storage system; assess the life expectancy or permanency of document images stored on a digital imaging system; and assess the capability of the digital imaging system to produce images that are faithful facsimiles of the input documents.

Document Selection: Identify and select representative samples of military service records for use. Ensure that the selected documents represent a complete set of the document characteristics that must be evaluated and addressed during the project.

Document Preparation: Determine what preparation and workload requirements are necessary for the complete and thorough input and conversion of documents; assess the capability of maintaining document integrity and control throughout the preparation, input and conversion stages; and, assess the capability of efficiently preparing documents for digital imaging input.

Document Input and Conversion: Establish the feasibility of converting paper documents to optical digital media; assess the system's capability to input and convert documents efficiently to digital form; assess the system's capability to provide image enhancement for representative samples of a full range of NARA documents; determine what levels and techniques of scan density and image enhancement work best for various types of documents; assess the human intervention requirements needed to perform document input and conversion; assess the manual and automatic system controls; determine what production methods are best to perform input and conversion in an operational environment; and, determine feasibility of scanning non-paper holdings such as roll microfilm and microfiche.

Document Indexing: Create an indexing system that has the dual independent capability of database information retrieval and document image retrieval; and assess the system's capability to add, modify and delete documents using the indexing scheme.

Verification and Quality Control: Assess the system's capability to provide and maintain a high level of quality control and accuracy of digital images during the input, conversion, indexing, storage, retrieval and output stages; determine how accurately the system inputs, converts, indexes, and stores the original documents; and, determine what techniques work best for maintaining high levels of quality control within the system.

Document Storage: Establish the feasibility of storing paper and non-paper document images on optical digital media; evaluate the storage capacity in terms of storage cost and efficiency; evaluate the system's capability to use mechanical devices effectively to retrieve automatically the optical media and send the requested document images to the requestor; and, determine the methods and requirements that work best in backing up the stored information.

Document Retrieval: Establish the feasibility of retrieving digital images of records instead of the actual physical paper or non-paper documents; evaluate remote index data retrieval; assess the retrieval capabilities of the system including speed of retrieval, intellectual control of the information, quality of retrieved images, and flexibility of retrieval requests; and determine staff and public reaction to the use of an image retrieval system as compared to use of original records or microfilm copies.

Document Output: Establish feasibility of creating hardcopy paper output from digitally stored data; and, assess the system output mediums to replicate and accurately provide facsimiles of the original documents.

System Operations: Evaluate the physical operation of the entire digital imaging system and assess hardware and software reliability; evaluate the capability to integrate various hardware and software components successfully into one workable system; and determine the relative costs and benefits of using an automated digital image system to support archival reference as compared to existing manual methods.

2.3 ODISS Design and Technical Requirements

ODISS was an NSZ research and experimentation project designed to evaluate how well digital imaging and optical disk technologies function with archival records. To perform this research it was necessary to acquire a research test system. NSZ, in consultation with the Office of the National Archives (NN) designed the system's technical requirements in order to examine the applicability of digital imaging technology to historical records.

For example, one major question is whether the technology could produce clear, legible images from NARA's textual documents containing a wide range of image quality problems. Therefore, the requirements specified scanners that could capture images at scan densities of 200, 300 or 400 pixels[46] per inch as well as performing "automatic and user controlled complex image enhancements."[47] These multiple capabilities were necessary to learn which combinations of scan densities and image enhancement techniques might produce useful images from the various NARA document holdings.

In another area, any large-scale conversion of NARA's massive quantities of records would need rapid throughput production. Since the decision was made to conduct a research project rather than an immediate full scale implementation, requirements were drafted about high speed production with archival documents. Thus, the requirements specified a high speed paper scanner with a rated speed of at least 40 images per minute in order to determine the "real" throughput rates for different archival record types.

The requirements in other areas such as the workstation characteristics and image storage required the contractor to provide state of the art equipment and capabilities. An effort to combine state of the art technology with a need to evaluate the feasibility of obtaining readable images from old NARA microfilm led to the requirement for a multi-format film scanner that could process various microforms including roll films, microfiche, and aperture cards.

Modeling the workflow helped define the specific actions that operators would have to perform during input and retrieval functions. This workflow also identified indexing and quality control functionality needed to perform these operations. The requirements for the

[46] Pixels, or picture elements, are discrete points (dots) sensed by the scanning equipment. Image sharpness is usually improved as the number of pixels per inch is increased.

[47] Image enhancement refers to the process of improving electronic image quality. Various enhancements are available; selection is usually based on original document characteristics and scanner performance.

information retrieval required the contractor to provide easy-to-use methods for conducting searches, obtaining images at the retrieval terminals, and printing hard copies.

In summary, the design process developed a model of an optical disk-based digital imaging system. It clarified and defined the hardware and software capabilities needed to conduct research experiments into the application of this new technology to archival work. The design effort set requirements for collecting data about NARA's optical digital image storage technology needs.

2.4 System Acquisition and Implementation Process

Federal government high technology procurements typically are complex processes following well defined, legally mandated administrative and procurement procedures. This policy has evolved to foster open competition among private industry bidding on government contracts, while ensuring that federal agencies acquire systems conforming to specifications. Some of the steps involved in agency procurements are: feasibility studies and requirements analysis, preparation of technical specifications, contract negotiation and award process, design reviews, factory acceptance tests, system delivery, and training. This cycle can follow various configurations, and may stretch out over several years. For example, ODISS was conceptualized in a NARA study almost four years prior to its actual on-site installation.

The Archival Research and Evaluation (NSZ), with assistance from Office of the National Archives (NN) staff prepared the ODISS functional and performance specifications. The specifications did not mandate any specific system approach or equipment. The government's basic requirements for system functionality were defined, and it was left up to the interested bidders to propose hardware, software, and operations design to meet the requirements. Since ODISS was such a complex, interrelated operation, the preparation of the government's specifications involved an integrated effort by NARA staff trained in different disciplines. To conduct the ODISS project, a NARA team was assembled to define system requirements designed to be issued in a Request For Proposal (RFP), conduct technical evaluations of the vendors' proposals, and monitor project implementation.

NSZ and NN put together a five member team to define the requirements for a digital imaging system. Two NSZ members provided knowledge of digital imaging technology as well as computer programming and ADP systems analyst backgrounds. Three NN staffers temporarily detailed to NSZ represented user needs, provided knowledge of military records and reference service operation for pension and compiled military service records, as well as general archival practices.

After the system's requirements were defined and an initial draft Request For Proposal (RFP) was completed, two of the NN members returned to their regular assignments, while the third was assigned permanently to NSZ to help further refine the draft RFP. A new staff member with experience in micrographics and automated production-oriented conversion systems subsequently replaced an NSZ member who pursued another NARA management opportunity. This left a basic three-person NSZ/ODISS project team to complete the RFP.

NSZ's three-member team became the lead group for implementing the project from contract award through system installation. The NSZ team's background experience included expertise in digital imaging technology, micrographics, automated conversion systems, and familiarity with archival practices and concerns. A similar staffing talent mix would be useful for any organization contemplating a digital imaging project.

Under the direction of NARA contracting officials, ODISS was not pursued as a research and development contract, but rather as a two-step, fixed-price acquisition procurement. The first step is used to screen prospective bidders for minimal technical competence, while the second step is a price bid by the technically qualified firms. The lowest price bid automatically wins the contract.

NARA released an Invitation For Bid (IFB) for ODISS in early November 1985, followed one month later by a pre-bidder's conference held in NARA's theater. Conducted by NARA's Contracting Officer, the bidder's conference addressed bidder's questions and/or concerns. Several vendors also submitted questions several weeks after the conference, requiring clarifications of the technical specifications from the Government.

In order to assure equitable evaluation of bidder's technical proposals, NARA convened a technical evaluation committee (TEC). TEC staffing consisted of the three NSZ staff members, rejoined by the two NN people who had worked on defining the system's requirements. These five committee members as a team determined which companies were technically qualified to build NARA's optical digital image storage system.

Each proposal was individually reviewed by a TEC member, followed by a committee meeting to achieve a scoring consensus. The committee's responsibility was to classify each proposal according to a pre-established set of criteria, and document the results. Evaluators numerically scored and ranked each proposal, using the weighing factors published in the solicitation document. When additional information/clarification was required from a bidder, the committee communicated with NARA's Contracting Officer. Meetings were also held with bidder's to obtain clarification of TEC member's questions.

The ODISS two-step process specified technical submissions by mid-February, 1986. NARA's contracting office received bids from seven vendors, with one bidder providing two technical proposals. Of the seven entries, three proposals were judged as technically qualified. The major part of the TEC committee's work was completed by early July, 1986. After completion of the technical review, the two NN staff members again returned to their original NARA duties.

The three technically qualified bidders were then requested to provide cost proposals. In this second stage, two out of the three qualified firms submitted cost bids. The bid opening was held in mid-August 1986, with the successful bidder being the lowest priced, technically qualified offeror, System Development Corporation (SDC) of Camarillo, CA. SDC was a relatively autonomous entity within the Burroughs Company. Following contract award, Burroughs and Sperry merged to form Unisys. SDC became a part of the new corporation, and the contractor was thereafter referred to as Unisys.

Prior to actually signing the contract, a government site visit to the Unisys facility in Camarillo, California was required. This visit verified SDC's in-house capabilities to perform the mandatory contract requirements successfully. The ODISS contract was officially awarded on September 8, 1986. The contract stipulated a one year duration for ODISS delivery and installation.

Project related activities began in earnest following contract award, for both NARA and Unisys. Now that the contract was in place, NARA had to identify a suitable ODISS installation site within the main NARA building. Several areas were examined prior to the final decision. Ongoing conferences between Unisys and NARA project staff focused on

facility and workflow issues. NARA staff also evaluated workstation furniture needed to support the ODISS equipment. The General Services Administration was responsible for preparing architectural and construction drawings for the ODISS room. These drawings and construction specifications were used in the award of a contract for ODISS site modifications.

In late October 1986, a systems requirements review (SRR) was held in Camarillo, California. This review was required as part of the original contract specifications, allowing Unisys to define and present a systems concept before formal development was started. The SRR was intended as an informal review process, with discussion of preliminary plans for software design, hardware configuration, a work schedule, and equipment installation and site plan information. This meeting was useful for Unisys to raise issues/questions about some of the mandatory requirements, and for NARA to evaluate Unisys's technical response to them.

A Critical Design Review (CDR) was held in mid-December 1986. This two day meeting was held at NARA to allow Unisys to present their understanding of the contract requirements and their planned technical approach to the ODISS project. The CDR also ensured that the detailed design solution and associated implementation plans and schedules satisfy the contract specifications. Project deliverables required at the CDR included a system functional description, a hardware description, and a software description.

Following the CDR, Unisys continued to work on the system development in Camarillo. NARA ODISS staff members were occupied with monitoring the facility construction, ordering ODISS workstation furniture, and reviewing Unisys technical submissions. Unfortunately, due to delays in equipment from subcontractors and other problems, Unisys was unable to deliver the system on time. Extensions of the delivery date had to be granted, and in late 1987, the government monitored the contractor's progress with a series of almost daily conference phone calls.

One of the contract clauses stated that prior to delivery, Unisys was required to demonstrate through a factory acceptance test (FAT) that the ODISS system met all technical requirements. In January 1988, Unisys notified the government that it was ready for the FAT. A NARA team composed of NSZ and NA officials travelled to the contractor's test site in Camarillo, California, on February 1-5, 1988. A factory acceptance test plan prepared by Unisys was used during the testing process, but due to system problems, Unisys failed to meet many essential requirements. Following this, the company was given more time for system development and checkout. A second factory acceptance test was held May 16-20, 1988. Although Unisys's performance was significantly improved in this second test, Unisys was again unsuccessful in meeting all of the NARA's technical requirements.

After the second factory acceptance test, it became clear that Unisys was unwilling to spend the additional time, money, and other resources necessary to satisfy all of NARA's technical requirements. The ODISS system at this point was deficient in page-to-page display and hardcopy printing speeds.

This left NARA with two choices. NARA could accept a reduced system from Unisys. Alternatively, it could find that the contract was defaulted and compel Unisys to pay for acquiring a system from another firm at some uncertain date in the future.

After much internal debate, NARA decided that although the Unisys system did not meet every original requirement, ODISS could still be used to conduct the research originally planned. In negotiations with Unisys, the government agreed to accept the lowered printing

speed capabilities, but exacted a price from the company. These costs also included direct costs for late system delivery and NARA FAT test expenditures. The method of payment was altered to make Unisys share more of the risk of a reduced system; two progressive levels of performance reliability (87% and 92%) were set that the installed system would have to meet over successive 30-day time periods before Unisys would receive full payment. Unisys had to accept a $175,000 reduction in the total contract price, which lowered the final contract cost to NARA about 18%.

The ODISS equipment finally was installed at the National Archives Building in July, 1988. A team of Unisys engineers and programmers also came to observe how the system performed, and they worked to correct many unanticipated difficulties that surfaced in the system's first production operations. Over the next few months, NARA and Unisys personnel often worked extended hours and weekends to solve problems that caused frequent system crashes or otherwise impeded system functionality. This effort paid off when the system was able to pass both levels of reliability tests. Unisys passed the 87 and 92 percent performance levels, and subsequently received the remaining contract funds due. By late November, 1988, the only Unisys person still on-site was the equipment field engineer technician, provided for one year under the contract.

ODISS subsequently ran smoothly in most respects most of the time, with Tennessee Cavalry CMSR conversion efforts continuing to progress under the direction of both NN and NSZ. CMSR conversion continued with NN staff performing the required operations activities. In spite of reliable system operation, a performance deficiency in throughput speeds became apparent.

ODISS operations were structured around file and block open/close parameters. For every operation, such as scanning, indexing and quality control, a file must opened, processed and then closed. File manipulations require computer system processing time. The original NARA estimate was fifteen images per CMSR file. As it turned out, the average file was closer to four images, with many reference cards containing only single images.

Unisys designed ODISS to work most efficiently when processing files containing multiple images. When presented with substantial numbers of files with only a single or a few images, production throughput delays were experienced. This was due to the operators having to wait for the system to complete routine file opening and closings. This deficiency was directly linked to the design of the System Manager and was unrelated to the scanning, digital imaging, or optical disk subsystems. The resultant lower than anticipated daily conversion throughput rates are not indicative of digital imaging systems in general. Corrective actions to improve the file processing cycle rates would have required modifications to ODISS.

The file conversion production shortfall raised concerns about the ability to complete the Tennessee holdings in the timeframe originally planned. It became obvious that the ODISS would require much longer to complete the conversion than had been previously expected. Although the ODISS production staff and the individual conversion equipment items were capable of faster speeds, they were collectively slowed down waiting for the computer system to service the file openings and closings.

In March 1989, Unisys, which was aware of the problem, submitted an unsolicited proposal with technical approaches for improving file processing. Following a series of discussions, NARA management from NN and NSZ decided in May 1989 that, since all testing goals

involving capture of the CMSR sample had already been achieved, cessation of the production operations would have no adverse effect on the project. The primary point in this determination was the fact that the Tennessee records were all very similar, and further conversion of that set of holdings would add little to the knowledge, experience, and statistics already accumulated. Consequently, CMSR conversion was terminated with completion of the Tennessee Cavalry records which contained approximately 54,000 files. The Tennessee Artillery and Infantry records were not converted.

The NN staff hired especially for ODISS was subsequently released to other duties throughout NARA. In following the project test plan, ODISS was used after that point for testing and accumulating data from the conversion of *ad hoc*, non-CMSR records from broader NARA holdings.

CHAPTER THREE

EXISTING NARA PROCESSES AND TECHNOLOGY UTILIZATION

3 EXISTING NARA PROCESSES AND TECHNOLOGY UTILIZATION

Paper has historically been the predominate medium for creation and retention of the federal government's textual records. The vast majority of NARA's holdings are paper records in various physical conditions. Because of preservation requirements, storage space limitations, and related costs, the government's use of microfilm has increased over time. In order to provide a better understanding of NARA's current use of paper and microforms, this chapter describes the historical growth and current applications of these two media.

3.1 Paper Records

For many centuries, paper has been the primary medium on which people have recorded information. Even today, despite the rapid growth of electronic information systems, paper remains the most common medium for storing and transmitting information. While some ADP enthusiasts prophesy the paperless office, automation typically has added more layers to an already well-papered world.

3.1.1 Physical Characteristics

The National Archives of the United States (NARA) is acutely aware of the prevalence of paper as the key information storage medium so far in human history. The historical records of the United States government preserved in the Archives are predominantly paper and, although only a small part of the huge quantity of papers generated by the large federal establishment, these archival holdings are voluminous. According to a 1985 study for preservation planning,[48] the records included over three billion pieces of paper. These 3 billion pages took up about 1.35 million cubic feet of storage space. The paper records in the Archives have increased steadily and by 1989 have grown to 4 billion sheets occupying 1,553,907 cubic feet.

A substantial portion of these paper records is in jeopardy of deterioration and eventual disintegration. The 1985 preservation study estimated that 160 million pages already have suffered major damage, 100 million pages are subject to damage by frequent use, and 270 million pages of 1940's to 1960's "quick copy" stencil, Mimeograph and Therma-Fax documents are deteriorating rapidly. The study concluded that 530 million pages are "at high risk of loss."

The 1985 study reported the results from a statistical survey of various characteristics of the Archives' paper holdings. The records include an estimated 950,000 bound volumes, which are 12.6% of the total volume in the survey. Two thirds of the paper is letter or legal size, about 10% is smaller and about 12% is larger. Many pages have more than one kind of imprint since 39% contain handwriting, 45% have typing, and 40% have printed text. About 36% of the pages have colored inks, and 8% are brittle. Only about 0.5% have faint images that are hard to reproduce, but this would still be about 15,000,000 pages out of the 3 billion in the Archives at the time of the survey. Fortunately, only 0.11% of the pages are so damaged that they have suffered actual loss of information.

---

[48] *National Archives And Records Service (NARS) Twenty Year Preservation Plan* (NBSIR 85-2999; issued January 1985).

3.1.2 Administration of Permanent Records

The basic mission of the National Archives is to preserve the federal government's records of continuing value and to make the records accessible to historians, genealogists, and other researchers. Records are made accessible through description and reference service. Description involves the preparation of various kinds of lists, inventories, and subject guides that help researchers find the records pertinent to their interests. Reference service means not only making original paper records available for examination at NARA facilities but also answering verbal and written inquiries about the records. It also involves providing copies of the paper records to researchers for fees.

The paper holdings of the Archives are under the jurisdictions of two major components - the Office of Presidential Libraries (NL) and the Office of the National Archives (NN). NN is much the larger in terms of both holdings and volume of reference service. For example, in fiscal year 1987, the nine presidential libraries under NL received 7,521 research inquiries and 10,425 researcher visits, while NN's Washington facilities and eleven field offices received 643,164 research inquiries and 206,645 researcher visits. In FY 1988, NL's 9,022 researcher inquiries and 12,233 daily researcher visits were much below NN's 514,083 researcher inquiries and 207,921 visits.[49]

The major preservation effort for paper records is the holdings maintenance program of the Office of the National Archives (NN). Recommended in the 1985 Archives twenty year preservation plan, holdings maintenance entails removing harmful fasteners, placing especially fragile documents in polyester jackets, and moving records into acid free folders and boxes. A laboratory in NN's Document Conservation Branch monitors the acid free quality of archival containers, conservators train staff in the proper preservation actions, and a new Holdings Maintenance Branch has been established to implement the program. In fiscal year 1987, more than 81,000 cubic feet received holdings maintenance action. In FY 1988, holdings maintenance actions were taken on 121,000 cubic feet.[50]

In addition, the NARA microfilming program produces microfilm publications of valuable records for both preservation and reference purposes. Microfilming can serve preservation goals by replacing the originals with film copies for researcher use and thereby saving the originals from the wear and tear of public handling. Microfilming serves reference goals by making multiple copies of the records available through purchase or use in the several NARA research rooms around the country. There are over 2000 NARA microfilm publications. In 1987, 40 more with 2,279 rolls of film were completed, and another 16 were completed in FY 1988.[51]

---

[49] *1987 Annual Report of the National Archives*, pp. 35, 100; and *1988 Annual Report of the National Archives*, pp. 31, 98.

[50] *1987 Annual Report of the National Archives*, p. 62, and *1988 Annual Report of the National Archives*, p. 59.

[51] *1987 Annual Report of the National Archives*, p. 35, and *1988 Annual Report of the National Archives*, pp. 31, 103.

3.1.3 Document Preservation and Conservation

NARA developed a "20 Year Preservation Plan" for textural records treatment based on studies conducted by the National Institute of Standards and Technology (NIST).[52] One study identified preservation needs and costs of current holdings, while a second study defined environmental standards for records storage conditions. Holdings maintenance is a key facet of the plan involving document flattening and placement in acid-free file folders and boxes, removing rusted staples and other fasteners, placing selected records in protective polyester sleeves, preservation copying, and treating weakened bindings. Although this costly process will meet the needs of many of NARA's holdings, many other historical documents will continue to deteriorate. This results from the sheer volume and advanced age of the documents which date from the beginnings of our government. Contributing to this problem is that over the years, many documents were mishandled and improperly stored prior to NARA's creation. Prevention of this continued deterioration is a major goal of NARA's five-level preservation program approach:

- Controlled environmental storage conditions
- Correct diagnosis and application of the most suitable archival conservation techniques
- Limited, careful document handling by researchers and conservator staff to minimize any further degradation
- Holdings maintenance activities
- Production of microforms of fragile and frequently requested documents

Microfilming and electrostatic photocopying technologies are used to recopy rapidly deteriorating documents. Maximum longevity can be achieved by storing documents under totally secluded conditions, where they remain undisturbed by researchers or staff. Extensive alterations to the National Archives building are required to meet the environmental standards outlined in the NIST studies. Storage without natural or artificial light in acid-free enclosures with the proper temperature and humidity conditions is also mandatory. Applying this approach to all documents in NARA custody would virtually eliminate records usage, which countermands NARA's policy of records accessibility.

For documents needing special conservation treatment, NARA maintains a conservation facility Time consuming and labor intensive techniques limit the quantity of documents receiving conservation treatment. In compliance with the 20 Year Plan, NARA augmented the Document Conservation Branch designed to treat deteriorating holdings. Additional NARA conservation staff possessing skills and diligence to undertake the meticulous tasks of document repair and reconstruction were hired. Also, modern conservation and analytical equipment was installed in NARA's laboratory. The professional staff also now ensures that documents receive optimum care during exhibit display at NARA and other institutions.

---

[52] At the time the "20 Year Preservation Plan" was developed, NIST was known as the National Bureau of Standards (NBS).

3.1.4 Retrieval And Finding Aids

Locating the right textual records for a researcher involves using a collection of tools called finding aids. These include general subject guides that cover the specific groups of records related to broad topics, preliminary inventories that give series descriptions for the records of a single federal agency, and box or shelf lists that give the physical locations of records in the archival storage rooms or "stacks." The guides and inventories usually have subject indexes, but the subject terms are often fairly general.

Some of the NARA prepared finding aids direct the researcher to the records of specific events. Most guide the searcher to the general location in the records where it is most likely that pertinent information can be found by the researcher making a box by box, file by file, page by page review. Very few finding aids to textual records take the researcher directly to the item level and the appropriate document. Such detailed item-level indexes typically are available only if they were developed by the government agency that created the records and then were transferred with the records to the Archives.

Work has begun to develop an archival database for retrieval of the records. The Office of the National Archives has defined requirements for an Archival Information System (AIS).[53] So far, however, automated reference is only a concept. A pilot to test the design concepts for AIS is being undertaken, but the outcome of the pilot lies somewhere in the future and full automation is even more distant. Meanwhile, the ODISS research test is the first implementation of automated indexing and retrieval of records at file level.

3.2 NARA Micrographics Policy and Operations

The National Archives has extensive experience with micrographics technology for document image storage. Dating from the early 1940's, NARA has microfilmed historic records for reference, distribution, and preservation. Years of production experience reinforce the fact that a quality microform product is a labor intensive process with inherent time, materials, and personnel costs. Large scale document conversions are significant financial commitments, and are affected by federal government budget constraints. Microform publications product sales and cost recovery potential are important NARA planning issues. Preservation microfilming, while still necessarily concerned with costs, employs time consuming operational techniques involving document conservation, precision image capture, film processing, and quality inspection criteria. Over the years, NARA has amassed a formidable collection of microforms through original filming and agency records accessions.

3.2.1 Evolution of Micrographics in the National Archives

The National Archives was a pioneer in federal government records microfilming. Even during NARA's formative years, the need for space reduction in records storage became apparent. World War II boosted microfilm applications due to increased records security concerns. Other government agencies began microform programs during this time, with many of their film products gradually acquired by the National Archives.

[53] Automation in the Office of the National Archives and the proposed features of AIS are described in NARA's 1988 annual report on pp. 55-56.

Archival application of micrographic technology is a prime NARA concern, due to the many documents required to be preserved forever. Problems with paper deterioration, storage conditions, physical handling abuse, and pilferage all contribute to the need for alternative preservation techniques. Silver halide microforms processed and stored under archival conditions[54] offer long term information storage. Original paper records can remain stored and undisturbed, while the information content is still available to interested researchers. Duplicate microforms are routinely distributed to NARA branches to service the information needs of researchers throughout the country.

3.2.2 Role in Records Storage and Preservation

NARA's storage of paper, film, photos, video, and maps is constantly expanding, as are the ever increasing access requests to this huge information data repository. In spite of retaining only approximately three percent of federal government records, NARA must accommodate increasing records holdings. Micrographics has historically offered greatly increased storage compaction over traditional paper based information systems. Microforms can save up to 98 percent of the physical storage space required by paper files. Once captured on microfilm, images are readily available for researcher use, and the rolls can be easily duplicated for information distribution to other user sites. Original records can be safely stored under archival conditions, protected from handling degradation, while the microforms are viewed and printed using suitable retrieval equipment.

Since the beginnings of NARA's microfilming efforts, NARA has amassed approximately one half million microfilm rolls, not including classified or accessioned microforms. This film repository has produced cost savings by reducing the storage space requirements, and facilitating a self-service type reference system for popular records series. Public and professional researchers are able to search records such as Civil War records, ship's passenger logs, Census records and many other popular genealogical holdings with minimal staff assistance. Permanent retention and preservation of NARA master microfilms is accomplished under carefully controlled security and environmental conditions. When additional film copies are needed, printing masters are used to produce the required copies.

3.2.3 Administrative Management

Administratively, microform production is under NARA's Preservation Policy and Services Division (NNP). The Special Media Preservation Branch (NNPS) oversees photography

---

[54] Standards for film processing:

ANSI IT9.1-1989, American National Standard for Imaging Media (Film), Silver-Gelatin Type; Specifications for Stability; 1989; American National Standards Institute, 1430 Broadway, New York, NY 10018

Standards for microfilm storage conditions:

ANSI PH1 43-1985, American National Standard for Photography (Film) Processed Safety Film Storage; 1985; American National Standards Institute, 1430 Broadway, New York, NY 10018

ANSI IT9.2-1988, American National Standard for Imaging Media. Photographic Processed Films, Plates, and Papers Filing Enclosures and Storage Containers, 1988, American National Standards Institute, 1430 Broadway, New York, NY 10018

(NNPS-P) and duplication sections (NNPS-D). NARA's original document microfilming, processing, and duplication functions are staffed with personnel especially trained in those operations. NARA's micrographics systems include cameras, film processing, inspection workstations, and silver film duplication. Roll film duplication is located at the NARA Annex on South Pickett Street in Alexandria, Virginia, and supports high-volume film duplication, film processing, and quality control inspection operations.

3.2.4 System Operations

NARA maintains an equipped and staffed in-house micrographics capability to handle a variety of original document filming and duplication requests. Original source document filming requires a series of interrelated administrative and production steps. A customer's microfilming request is received and logged in, followed by an archives technician pulling the appropriate records holdings. The documents and order form are submitted to the microfilm section, where the job is logged into the service order system. After the documents are provided to the camera area supervisor, equipment and film formats are selected based on customer requirements. A microform camera technician operates the camera system according to prescribed procedures, feeding the documents one by one until the batch is completed. The exposed, undeveloped microfilm is removed from the camera, and sent to the film lab for developing and technical quality control inspection. Following this step, the developed film is returned to the camera area for inspection. Any defective imagery is noted, and documents reshot as required for splicing.

NARA procedures require the production laboratory to inspect film products for technical qualities, while information content verification is the responsibility of the custodial unit. Microfilm copies are printed and subsequently developed as required. Microfiche production conforms with these operational steps, with the added procedure of cutting the 105mm rolls into individual microfiche following processing and duplication. After labelling and boxing the film(s), the request is logged out of the camera area tracking systems. The requesting custodial unit is notified to retrieve the original documents and completed microfilm. The custodial unit's archives technicians and supervisory archivists inspect the order for quality and completeness. For outside billable requests, orders are processed for payment and packaged for customer delivery.

Customer requests for film duplications follow similar paths, except that instead of pulling original documents that require microfilming, print film masters or original camera films are retrieved. The use of printing masters is preferable, since the original camera master films can remain in secure storage. Depending on microfilm format and request volume, the masters are duplicated in either NARA's camera master processing lab or at the Pickett Street facility. Microfiche duplication is performed in the main building's film processing lab, since this facility has equipment to handle the wider film format.

The Pickett Street facility duplicates 16mm and 35mm roll films. Masters are duplicated on positive or negative silver halide materials. The duplicates are quality checked and delivered to the custodial units for order fulfillment.

3.2.4.1 Camera Area Operations and Production Statistics

Microfilming throughput rates are affected by document characteristics. NARA owns several high-speed mechanized transport cameras, but they were determined to be hazardous for many of NARA's aged records holdings. Fragile, deteriorating documents and bound volumes

are not suitable for the high speed microfilmers. NARA determined that these mechanized camera systems are most useful for index and similar cards which are printed on durable paper stock, or more modern 8.5" x 11" inch standardized office documents. Most NARA documents are microfilmed using manual feed planetary camera systems.

Although this has not always been the case, microfilm production is now under NARA's production standards program. The camera area supervisor assigns a microfilm production job to a camera operator. Production rates are based on the varied document physical characteristics and the camera type. Operator rates range from a low of 994 images up to 2,331 images per day for flat work.[55] This variation is due to document conditions and needs for special handling. The lower rate is for documents individually filmed under a glass platen. The higher rate is for totally prepared documents in good physical condition.

3.2.4.2 Film Processing Operations and Production Statistics

NARA uses silver halide materials for both microfilming and film duplication. Silver films require carefully controlled chemical development to create consistently high quality, permanent images. NARA has table top and deep tank processing equipment for low and high volume film throughput speeds. Exposed microfilms are delivered to the processing lab, where a trained technician is responsible for operating and maintaining the equipment. This station requires careful monitoring of processing speeds, temperatures, and chemical solutions. Each roll has a leader attached, followed by automatic film travel through the processor's transport system. NARA owns a tabletop processor for 16mm and 35mm films, and a larger floor-standing unit for 105mm film widths. Each developed roll is examined for density, contrast, resolution (sharpness), image placement, and physical defects such as scratches and other visible problems.

As required, the processing lab also produces all silver microfiche duplicates, and occasional roll film duplicates. Automated printers expose silver direct and negative film duplicate materials. These print films require more frequent processing chemical changes due to the residue buildup in the development solutions.

The camera master processing station operates at the equipment rated speed. For example, the table top processor runs at 10 feet per minute. Station throughput is affected by chemical preparation, equipment warm-up, calibration, and area clean-up activities. Daily film processing rates are 24 rolls of 35mm film, and 39 rolls of 105mm film, which may be accomplished concurrently. The processing technician also inspects 16mm and 35mm microfilms at 20 rolls per day, and 24 (50 feet capacity) rolls of uncut microfiche per day.[56] Silver duplicate microfiche production, under the NNPS production program, requires a daily minimum of 702 cut fiche, and 78 rolls of 105mm film duplicated per day.

3.2.4.3 Quality Control Operations

Monitoring product quality is a vital part of microform systems. Microform production requires maintaining precise tolerances in order to conform with industry guidelines and Federal Property Management Regulations (FPMR). Experience has shown that microform

[55] Production rates from NNPS Production Standards guidelines.

[56] Production rates from NNPS Production Standards guidelines.

systems which fail to implement an adequate quality assurance program suffer inferior image quality and loss of vital document information.

NARA's microform inspection procedures combine technical and content verification. The processing technician performs a technical inspection of the processed microforms immediately following film development. NARA's micrographic technician uses precision measurement tools to evaluate the film products for exposure, development, and image quality. This permits the technician to examine not only individual film rolls, but also to monitor camera station equipment and operations. This technician observes trends in film output qualities, and makes recommendations regarding corrective actions when problems occur. Quality problems can include incorrectly exposed or blurry images, erroneous camera reduction settings, and physical defects such as film scratches.

The custodial unit is ultimately responsible for conducting informational content verification of microfilm or microfiche. *NARA Microfilm Publication Procedures NN 88-01* specifies the extent of inspection required, which ranges from one hundred percent image-by-image comparison against original paper documents, down to sampling of images with reference to paper documents only to resolve questions. The specific verification level is determined by the archivist in charge, and is typically specified in the filming instructions for the records series. Image rejection will result in pulling and refilming affected pages. The refilmings are then developed, and spliced into the master rolls in correct chronological order. Defective microfiche are usually replaced as no suitable procedure exists to replace individual imagery. Quality control must be applied to all system production steps, including microfilm duplicates and hardcopy output.

NNPS has a published set of quality standards, which define the major categories of filming errors. This document specifies the number of errors per roll allowed, which currently results in an estimated two percent rejection rate.[57]

3.2.4.4 Testing and Storage Requirements

NARA produces camera master microforms in conformance with FPMR requirements. Methylene Blue chemical analysis precisely determines the amount of residual thiosulfate in film emulsions, which correlates to the archival life of silver halide microforms. Archival testing is done in the Research and Testing Laboratory (Room B-3) twice weekly, with films being rewashed and retested when they do not meet specifications. NARA's film processing technicians maintain logbooks with detailed test results, augmented with processing chemical and water supply data.

Endurance of micrographic quality depends on several factors, including the film's chemical stability, and processing and storage conditions. Film emulsion stability is determined during the manufacturing process, while developing and storage conditions are under user control. User copies exposed to continual handling suffer from dirt, abrasion, fingerprints, and contamination with foreign matter. Due to these conditions, film copies such as those maintained in NARA's Microfilm Reading Room cannot be considered candidates for long-term preservation. Microfilm intended for permanent preservation requires proper processing, minimal handling, and appropriate storage. Archival microforms must also be

---

[57] Estimated NN image reject rate as of March 10, 1989.

provided with fire protection, water protection, humidity control to eliminate fungus growth, elimination of atmospheric contamination, and theft protection.

Guidelines for long term storage[58] include:

- Less than 0.014 micrograms of residual thiosulfate per square centimeter remaining in the film emulsion

- Temperature below 68 degrees, relative humidity of 15%-40% for cellulose base silver films

- Air conditioning with positive air flow, free from airborne gases, dirt, and other contaminants

Although archival microfilms should be kept under these conditions, reference films have more environmental flexibility. An ongoing film inspection program, based on a statistical sampling plan, is also recommended for permanently stored films.

NARA utilizes the facilities of National Underground Storage, Incorporated located at Boyers, Pennsylvania for long-term storage of master silver halide camera negatives. Boyers is a rural area of western Pennsylvania, approximately 55 miles north of Pittsburgh. Situated 220 feet beneath the surface of a mountain in what was formerly an abandoned limestone cave, the facility consists of huge rooms available for records storage. Each is sealed off from the tunnels and naturally maintains a constant 55 F year-round. Dehumidifiers keep the relative humidity to the levels specified by customers. NARA's microfilm storage room, for example, is maintained at 30%.

3.2.4.5 Duplication Operations

NARA camera masters are stored under archival conditions, while users are provided durable, but easily replaceable microform copies. Depending on customer requirements and original film format, the duplicates can be positive or negative polarity in 16mm or 35mm rolls, and 105mm microfiche. NARA's Special Media Preservation Branch (NNPS-D), which operates at the Pickett Street Annex, generates over 10 million linear feet of duplicate microfilm each year. New processing equipment was installed to help with these massive duplication requirements. NNPS-D's printers operate in a roll-to-roll mode using large-capacity master and print material supply reels.

NARA's duplication workload fluctuates based on customer requests, in addition to ongoing large holdings conversions. Camera master negatives are retrieved from Boyers when a printing master does not exist. Since this operation puts the films at risk, NARA staff creates printing masters of unduplicated films as time permits.

3.2.4.6 Production Problems

Any large scale production operation which involves complex, precision equipment is susceptible to a variety of systemic and/or routine problems. Micrographic systems are no exception, and in spite of the best efforts many systems operate with inherent deficiencies.

---

[58] Refer to Footnote 54.

These can involve areas such as: personnel, operations, equipment, facility, and product quality. NARA also is susceptible to problems, examples of which include:

- Interrupted workflow patterns due to absence of integrated production workspaces
- Need for upgraded and/or replacement microfilm equipment, some of which is antiquated
- Requirements for increased controls over facility power, water, temperature, and humidity
- Improvements to the microform supply cold storage area
- Need for creating print masters for all camera rolls stored in Boyers, Pennsylvania

NARA intends to correct these deficiencies in the new Archives II facility. NARA staff continue to produce quality microfilm products which meet applicable specifications in spite of these obstacles.

3.2.4.7 Document Handling Considerations During Conversions

NARA uses microfilm to protect fragile high-use and intrinsically valuable documents from repeated reference handling damage. Document conversions are conducted with minimal damage to the original records. A close working relationship between the conservation lab and the production facility is vital when planning a major conversion project. Holdings preservation is necessarily concerned with careful handling of brittle, aged documents, since it is difficult to justify any conversion project if the process itself causes document deterioration. This is an ongoing dilemma for both microfilm and digital imaging systems. The capture station equipment has to be gentle with archival documents, while maintaining a reasonable throughput rate. Automated document handling systems have evolved over the years, in attempts to achieve true high speed processing with negligible document wear and tear.

Although it may not be possible to manufacture a microform camera or digital scanner which can absolutely guarantee that no document will ever be damaged, degradation can be minimized if some degree of caution is exercised by equipment operators. Documents which appear to be excessively delicate, or of unusual size or binding characteristics, should be captured at a low speed station.

3.2.5 Information Retrieval from Microforms

The preceding sections focus on microform production within NARA. An entirely different aspect is actual microform utilization. Information retrieval using NARA microforms involves: searching the available index data to identify information location; locating the desired image in the microform repository; and, image viewing and/or printing using NARA's microform retrieval equipment. The following sections describe NARA's microform reference operations and data retrieval considerations.

3.2.5.1 Utilization for Research

This section describes the typical steps facing a researcher interested in using the microform holdings in the main National Archives building. There are several broad categories of user groups, ranging from professional researchers performing client searches and academic researchers conducting scholarly analysis to novice genealogists just beginning to learn how to trace their progenitors. Since researcher skill levels vary widely, NARA's staff and operational procedures are organized to support the needs of these diverse groups. A major NARA mission objective is to maintain and make available to researchers the permanently valuable records of the federal government. These records were originally accumulated during the normal course of government business, and were not specifically created to aid users searching for ancestral information. Original paper records may be provided when microfilm copies are not currently available. NARA maintains two facilities in the main building: a Microform Reading Room on the fourth floor, and a Central Research Room on the second floor.

The Microform Reading Room contains more than a hundred thousand microfilm rolls, and the Central Research Room allows access to original records retrieved by Archives staff. It should be noted that the National Archives has other research facilities: eleven regional archives branches strategically placed throughout the country, fourteen records centers, and eight presidential libraries. NARA recently broke ground for a new Archives building on the University of Maryland campus in College Park, Maryland.

A typical search day begins at the front desk of the Pennsylvania Avenue entrance. Researchers must sign in at the guard's station, and undergo a security search of their hand carried items. Researchers who require an identification card are directed to the second floor, where a NARA staff member performs the identification and verification process. Researchers planning on using only microform records may proceed unescorted to the fourth floor research area, where a continuing audio visual presentation is available to visitors. This show provides a brief introduction into the National Archives, its genealogical holdings, and how to proceed with records utilization. Visitors can also discuss their information needs with NARA volunteer staff aides in that area.

Researchers log in upon entering the Microfilm Reading Room. NARA staff are available to describe available microform holdings and provide instruction in room procedures and retrieval equipment operations. Since this facility is primarily a self-service operation, the researchers at this point are generally on their own. The vast majority of microforms in the Reading Room are 16mm and 35mm roll films, wound on plastic reels and stored in protective cardboard boxes. Limited search aids exist for some of the filmed records, while others require manual search efforts based on all known search criteria. With the microform identification information in hand, the researcher then retrieves the required microfilm if that film roll is correctly filed and not in use.

The user then selects a film viewer on a first-come, first-served availability basis. NARA's roll film viewers are manual hand cranked models, requiring film threading and adjustments for image focus. A researcher would typically wind slowly through a film roll, stopping occasionally to determine the proximity to their desired image. Once the target images are located, researchers carefully peruse the images to determine if it contains the desired data. If no other images of interest are contained on that roll, then manual film rewinding onto the supply reel is required. The researchers are responsible for returning the reboxed films to the correct storage cabinets.

If needed, paper copy prints from microfilms are available to researchers by purchasing a "fare card" from an automated dispenser. The researcher proceeds with the microfilm and fare card to one of several viewer printers in the Reading Room. Prints are produced on demand at the push of a button, followed by rewinding and return of the microform to its storage location. When finished, the researcher returns to the guards' desk at the main entrance on the ground floor for a search of personal belongings prior to exiting the building. See Figure 3-1 for a graphic illustration of the researcher activity workflow.

Research can be time-consuming, requiring patience to deal effectively with the existing procedures. Development of effective search technique skills requires practice and hands-on experience. Not all search sessions are successful, and many times result in discovering additional search avenues to investigate rather than obtaining the desired complete answers. Depending on the extent of the search and information needed, a researcher may examine several large record holdings to gather a more complete picture. For example, ships passenger lists are useful in verifying when an individual or family arrived in a major American port city. Census records are useful for determining more detailed information about housing and family member issues. Unfortunately, a researcher is not able to go to one single NARA index source or computerized workstation search system. Much of the search time is spent in trial and error. Indexes and other finding aids for series relevant to genealogists are often incomplete, and researchers themselves frequently possess only limited personal information on which to base searches. Repeated searches are often required. The most successful researchers are typically those who already know a great deal about the search topic of interest.

3.2.5.2 Image Quality Considerations

Image quality for much of the microfilm in the Reading Room is marginal. Due to the age and extreme high use, many rolls contain excessive scratches, blurry images, low contrast, and other problems. Many of the master films were created prior to installation of modern production equipment, and poor document quality also contributed to the challenging microfilm task. This image quality problem makes the typical researcher's job more difficult, as more time is required to decipher the image content. Hardcopy prints produced from the films are frequently of less then optimum legibility.

3.3 CMSR Reference

This section describes the existing reference activities for the compiled military service records.

3.3.1 Reference Activity

Reference service for the compiled military service records falls under NN's General Reference Service Branch (NNRG), and it is performed by the Pension and Military Service Records Section (NNRG - P).[59] This branch handles both mail-in inquiries and requests from visitors to the National Archives Building. While requests from public visitors approximate 300 to 400 per day, detailed staff productivity statistics are not maintained about reference work for the walk-in public.

[59] The discussion of the CMSR reference activity is based on data provided by Tod J. Butler, Chief of the General Reference Service Branch (NNRG), during a meeting on June 15, 1989.

Microfilm Research Process

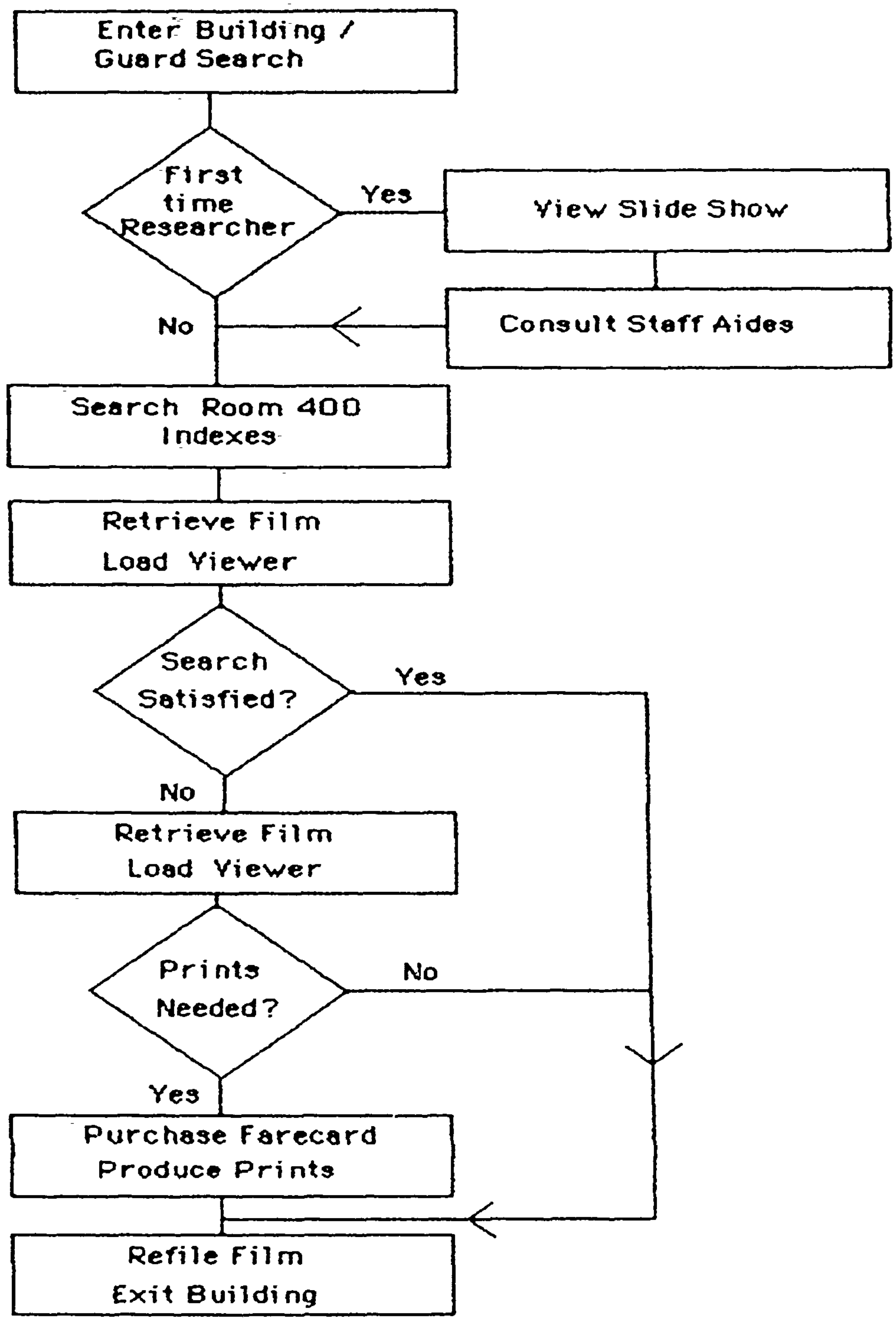


Figure 3-1

However, the staff performing reference work for the inquiries received by mail must meet production standards and their pay is based on their production rates. Because the work is assigned in standard batches and the time spent on each batch is monitored, detailed statistics are available for reference work to answer mail-in CMSR inquiries.

3.3.1.1 Staff and Organization

The CMSR reference work is divided between three groups. Within NNRG, one group of archives technicians performs searches to answer the inquiries. If the search is successful and a file is located, the technicians pull the file and turn the package over to a second group. This second group makes copies of the file. The inquirer is notified by mail that a file has been located and that, after NARA receives payment, copies of the most significant documents in the file will be sent. The third group is the mail room staff that keeps the copies until notified of NARA's receipt of payment and then mails the copies to the requestor. A substantial number of copies are never claimed with a resultant loss of staff and supply costs to NARA. A flow chart of the CMSR mail-in research process is shown in Figure 3-2.

3.3.1.2 Walk-in Public Reference

In addition to the mail-in CMSR requests, researchers come to the National Archives building to search the CMSR records. Some of the records have been microfilmed and researchers must use the film of these records, which is available in the Microfilm Reading Room. Other CMSR records have not been filmed, and researchers are provided with these original paper files in the Central Research Room on the second floor.

Researchers generally start their searches for CMSR files by reviewing microfilmed indexes. In some cases when the indexes lead to microfilmed records, the researcher uses the microfilm in the Microfilm Reading Room. Although NARA staff provides some assistance, this microfilm research is chiefly a self-service operation. Prints can be made by the researchers on reader-printers, which are activated by fee cards bought by the researchers. When the index search leads to CMSR files that have not been filmed, researchers fill out a request slip. These are collected periodically by NARA staff, who locate the files and bring them to the Central Research Room. The researchers examine the files and, if needed, make copies at a self-service copier for a fee. The walk-in CMSR research activity is diagramed in Figure 3-3.

CMSR Mail-in Search Process

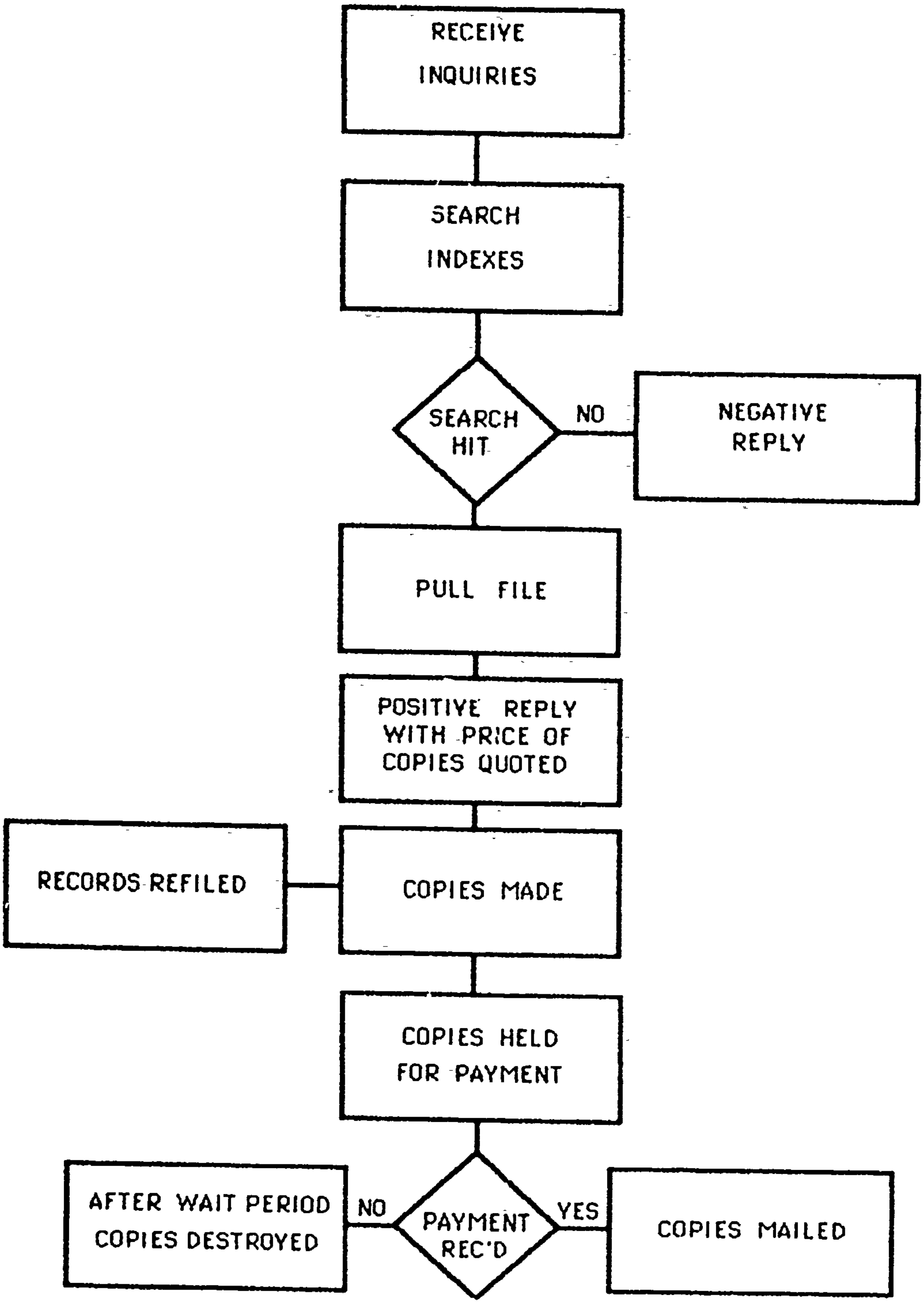


Figure 3-2

CMSR Walk-in Service

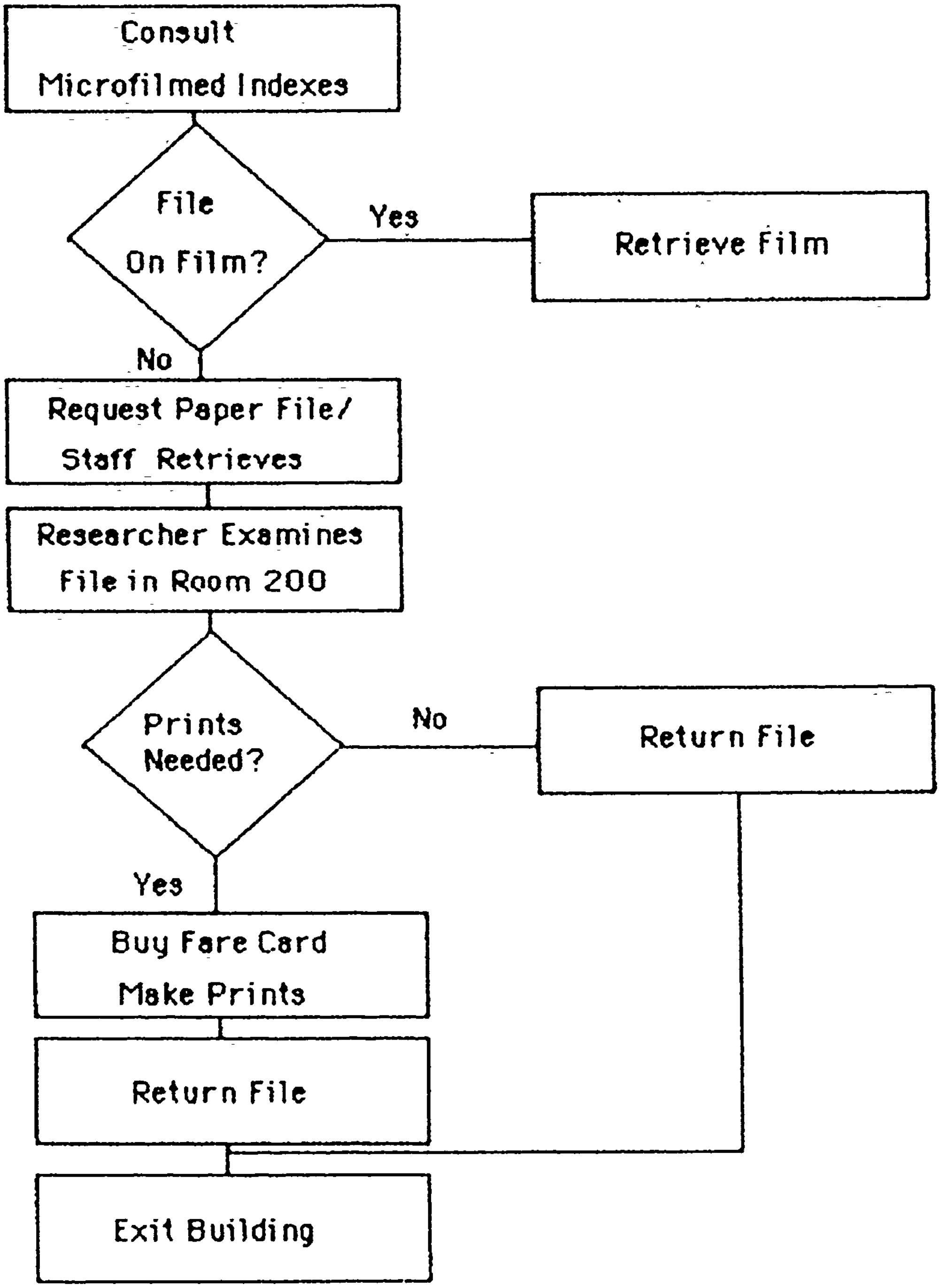


Figure 3-3

CHAPTER FOUR

ODISS SUBSYSTEM DESCRIPTIONS

4 ODISS SUBSYSTEM DESCRIPTIONS

The Optical Digital Image Storage System (ODISS) was installed at the National Archives in July 1988. This chapter of the report provides a basic introduction to the ODISS equipment configuration. For more detailed descriptions of the system equipment, software, and operating procedures, refer to Appendix B on page 206. Photographs of the various system components are presented in Appendix H which begins on page 360.

4.1 General System Concept

ODISS is a research test facility. The purpose behind its acquisition was to test the suitability of digital image and optical disk technologies for the conversion, storage and retrieval of archival materials. Based upon research test requirements, ODISS was designed as three functional or production subsystems: conversion, storage and retrieval. The conversion subsystem is responsible for creating, from a source document file, an indexed temporary magnetic disk file of digital images. The conversion subsystem thus includes image capture, indexing, and quality control, with the subsequent recapture and replacement of defective images. The storage subsystem is responsible for the transfer of completed image files from magnetic to optical disk. Within the digital imaging industry, this process is generally called "archiving." The retrieval subsystem is responsible for database reference and retrieval of archived images. The retrieval subsystem thus supports query of the image file index, screen display of query results, retrieval and screen display of images corresponding to the query, and print-to-paper capability for both images and query results.

4.1.1 File Data Structure

Tennessee Confederate Combined Military Service Records (CMSR) were selected as the test set of records for the primary ODISS conversion production test. Tennessee CMSR records are arranged by regiment and company and thereunder by individual. All records corresponding to an individual's service in a particular company were originally filed in jackets housed in Hollinger boxes. The arrangement of the CMSR source documents can thus be summarized as follows: documents corresponding to an individual's service in a regiment are filed in jackets which are stored in Hollinger boxes. The storage of captured images in ODISS follows a parallel structure. Scanned images of documents are stored as data files of individual service, and these files are grouped into blocks, corresponding roughly to the contents of one Hollinger box. Like the corresponding source documents, ODISS Tennessee CMSR images are indexed at the file level used to control storage and retrieval of images.

Also as is the case with the corresponding source documents, the ODISS Tennessee CMSR index is maintained as a separate data file from the stored images. Completed images are stored as image data on optical disk. Completed index entries are stored in database format on magnetic disk, using Unify relational database software. Principal database entries for each ODISS Tennessee CMSR file which control its retrieval include the name, company and regiment of the serviceman, as well as the location on optical disk of the image file of the service record.

4.1.2 Conversion

Conversion begins with source document preparation. Primary image capture by a high-speed scanner follows. Captured images are indexed for later reference. Images are reviewed for quality and the index for accuracy. Indexing errors are corrected immediately during the

quality review process, and problem documents and the corresponding images are flagged for rescanning. Rescanning of problem documents is accomplished by a low-speed scanner capable of producing improved quality with problem documents, at a sacrifice in the total time (i.e., throughput) required to scan the document and write the digital image to electronic storage. Acceptable images are substituted for defective images in the serviceman's electronic file. During the conversion phase of processing, index data and all scanned images are stored on magnetic disk.

4.1.3 Storage

The ODISS long-term storage system uses two-sided write once, read many times (WORM) digital optical disks, stored in an autochanger (jukebox). WORM optical disks are so named because they are not erasable. When blocks of scanned images are ready for long-term storage, the images are written to optical disk. The first images are written to side A of the first optical disk. When side A is full, it is copied to the first side of a second disk, creating the backup copy. The backup copy is available for immediate reference use, while the first disk is flipped to side B, and the archiving, or writing from magnetic to optical disk, of completed images continues. When both sides of the first disk are full, its side B is copied to side B of the backup disk, and the backup disk is stored in an alternate location. The first disk becomes the primary retrieval disk.

4.1.4 Retrieval

The ODISS retrieval subsystem consists of two staff workstations and one public workstation[60]. The workstations display both image and textual data. The latter consists of instructions for searching the index database and displaying information retrieved from it. A search of the index begins with entering information into any of the thirteen CMSR search fields that include last name, first name, middle name, and code values (from displayed tables), rank in, rank out, regiment, and company. Where information is not known, the field is left blank and the search is made using only those fields for which information is entered. When a file matches the search query, the file control number, index information for all the fields, and the number of images in the file are displayed on the screen. When there is no match, a message on the screen conveys this information. If several files match the search query, all of the results are displayed.

File images are retrieved by executing a function key. Other function keys rotate the image and enlarge the image through a zoom capability.

4.1.5 Duties of the ODISS System Manager

The ODISS operations staff included a system manager responsible for ensuring that all system initiation and monitoring functions were coordinated smoothly from the initiate and monitor subsystem. Both the personnel function and the system function came to be called the "system manager." The personnel system management function is related most closely to the operation of the three terminals, System Manager (see 4.2.5.1), CSE/ARS (4.2.5.2), and Archive Control (4.2.5.3). These three terminals are located at the system manager's station. The person designated as the system manager is responsible for additions, deletions,

[60] Any image workstation (except rescan) is capable of serving interchangeably as an index, QC, or retrieval workstation.

inquiries, and modifications to various files on the system manager database. The system manager could also delete files from magnetic disk at the Capture Server Element terminal or from optical disk at the Archive Control terminal. The system manager periodically consults information available from the three system manager terminals to determine workflow through the conversion and storage system.

Initiation of standard programs and production of routine reports are accomplished from the system management terminal. Disk diagnosis is accomplished from the Archive Control terminal. The system manager maintains a manual Archive Block Status Log, listing the status of all blocks between completion at the rescan station and final archiving to optical disk. The system manager maintains all accounts of file and block deletions from magnetic disk, file deletions from optical disk, name list data for partial blocks entered at the rescan station, and test results from disk diagnosis. The system manager has charge of the final preparatory acts prior to an archiving of image files. The most important responsibility of the system manager is the initiation of the final archiving to optical disk. The system manager creates duplicate copies of optical disks, has general responsibility for the mounting and dismounting of optical disks, and does database backup daily, initiating backup from the system manager terminal.

4.2 Hardware and Software Configuration

4.2.1 Major Subsystems

Conversion, storage and retrieval are accomplished in ODISS by a linked hardware configuration. This configuration includes the following five hardware subsystems: capture, workstation, archive, print, and initiate and monitor (system manager). The capture subsystem is responsible for the scanning of documents, the temporary storage of document images, and the storage of index data. The workstation subsystem is responsible for indexing, quality control and retrieval of documents. The archive subsystem is responsible for the writing of completed blocks of images to optical disk. The print subsystem is responsible for producing paper copies of images and query results. The initiate and monitor subsystem is responsible for the overall control of production and hardware within ODISS. This subsystem acts as the overall system manager.

Each subsystem consists of hardware specific to its task, supported by dedicated file servers and, with the exception of the archive subsystem, magnetic disk storage. Servers are microprocessors which are primarily responsible for coordinating data flow between large-scale magnetic disk storage and subsystem hardware components. System management software also resides on a server, and system management functions are controlled through a server. Communications between each subsystem's hardware components, its servers and the system manager are handled on hard wires. Transmission of image data requires RS422 cable; transmission of standard character data from database and other system software requires RS232 cable. Communications between servers are handled with a multibus interconnect; communication between the archive subsystem server and the archive subsystem is handled via a Small Computer System Interface (SCSI) bus.

For the ODISS configuration, Unisys-supplied C-language software for the control of workstation activities, such as the menus for indexing and the display of code tables. Unisys's software provides the links between system components and coordinates workstation activities with other software modules such as the printer module or the system manager module.

ODISS utilizes three operating systems. The workstations run under MS-DOS, the Heurikon servers run under VRTX, and ODISS as an integrated system runs under UNIX. The Unify relational database management system used to index Tennessee CMSR records also runs under UNIX. Links between components and coordination between different operating systems are provided by Unisys software. Most of this software is written in the C language, supplemented by assembly language routines where needed to maximize performance.

4.2.2 Digital Image Scanners

ODISS has four scanners, three for paper documents and a fourth for microforms. All ODISS document scanners are capable of some level of image enhancement.

4.2.2.1 High Speed Scanner

ODISS employs a Photomatrix high-speed scanner, capable of scanning both sides of a document at a rate in excess of 20 documents a minute. The scanner is capable of scanning documents at 200 dots per inch. It has two components, a scanner transport unit and an electronics unit, each in a separate enclosure. Also included with the scanner is a high-resolution monitor, which displays images as soon as they are captured by the scanner, providing the operator with immediate feedback about the quality of the scan.

The high-speed scanner uses a two-belt vacuum hold-down transport system to scan both sides of a document in one pass through the scanner. The scanner converts light reflected from the document into a raster map. The scanner electronics unit assigns a binary value to each pixel, compresses the raster map, and outputs the compressed image to magnetic disk storage.

Scanner operation is developed around the file block concept. Each block contains one or more files of documents which are stored and controlled by block number, and by file number within each block. Scanning begins only after obtaining a block number for the group of files to be scanned. The block is opened, images are scanned and stored in files within the block, and at the close of scanning the block is closed. Scanner operators also control the opening and closing of files within blocks. In the Tennessee CMSR conversion, the high-speed scanner served as the primary conversion scanner. All images corresponding to any one individual's service record were placed in one [separate] file.

4.2.2.2 Low Speed Paper Scanners

ODISS has two low-speed paper scanners which operate without a paper transport, making them ideal for documents too fragile to scan in the high-speed scanner. Each low-speed scanner has a flat glass platen on which the original documents are placed. One low-speed scanner is a binary scanner; the other is a gray scale scanner. The binary scanner produces an image in which all pixels are assigned one of two values: black or white. The gray scale scanner produces an image in which pixels are assigned one of 256 values, ranging from pure white to pure black. The binary scanner uses hardware image enhancement; the gray scale scanner uses software image enhancement.

4.2.2.2.1 Binary Scanner

The binary scanner is capable of scanning documents up to 11" by 17". It scans at 200, 300 or 400 dots per inch (dpi). This scanner employs three image enhancement modes: character mode, photograph mode, or character/photograph mode. Character mode clarifies the image by brightening the light areas and shading the dark areas further. Photograph mode increases the clarity of documents containing halftones, i.e., a significant amount of visual information which is neither at the very bright nor at the very dark end of the scale. Character/photograph mode is suitable for documents which contain both character and halftone information. The binary scanner also used a beta-test version of the Image Processing Technologies image enhancement device called the Scan Optimizer.

The scanner is controlled from a workstation which consists of a 286-based personal computer with keyboard and monitor. In the CMSR conversion, this scanner was used to rescan documents when the high-speed scanner had not produced an acceptable image and to scan documents too fragile or too large for the high-speed scanner. The workstation is capable of scanning images, storing scanned images on magnetic disk, and receiving images for storage from the gray scale scanner.

4.2.2.2.2 Gray Scale Scanner

The gray scale scanner captures document images with 8-bit gray scale (for a total of 256 values) at 200, 300 or 400 dots per inch (dpi). It is capable of scanning documents up to 11" by 14" in size. Scanner hardware provides a raw image as unprocessed data converted directly from the analog CCD output into 8-bit digital form.

The gray scale scanner is controlled from the same workstation as the binary low-speed scanner, but has a dedicated high-resolution monitor. Software installed at the workstation performs image enhancement on the raw image received from the scanner. The image enhancement terminal can send the enhanced image directly to the high-resolution monitor for display, or the enhanced image can be binarized and bit-packed (stored so that each byte represents 8 pixels rather than one) so that it can be displayed on a low-resolution monitor or transferred to capture subsystem disk storage. Enhancement software allows image enhancement techniques to be applied across the whole image or within a region of interest anywhere within the image designated by the operator before enhancement techniques are applied.

4.2.2.3 Multi-Format Microform Scanner

The multiformat microform scanner is capable of scanning microfiche, aperture cards and 16mm and 35mm microfilm. The scanner consists of two units, the scanning component and the electronic image processing component. Each film medium requires a mechanical adapter to position and hold the microform in place. For reel film, input and take-up reels are provided. The operator is provided with the capability to fine-position the image via keyboard input on the scanner's host computer. The microform scanner is capable of the equivalent of 400 dpi resolution on an 8.5" by 11" document which has been reduced by 48x. Processing hardware used in the microform scanner is similar to that used in the high speed scanner.

4.2.3 Workstation Subsystem

The same hardware configuration is used for indexing, quality control, demonstration, and staff and public retrieval. ODISS contains eight workstations, any of which can perform any system information processing function. The initial system plan called for two workstations to be assigned to indexing in the processing of CMSR records, and two to quality control. Two workstations are available to support retrieval in the main ODISS installation, and one workstation each is available to staff and to the public for retrieval. All ODISS workstations consist of 286-based personal computers, equipped with image processing boards and 19-inch black and white video display monitors. During use, workstation display screens are split between an image display area on the left side and an alphanumeric display area on the right. Images are shown at 150 dpi, with a capability to display partial images at the original scan resolution of 200 dpi. A 2x zoom is also available. Documents up to 8.5 x 11 inches can be displayed at full size.

The core server cabinet contains three Heurikon HK68/M10 single-board computers dedicated to the workstation subsystem. These are based on the Motorola 68010 microprocessor, with a 10 Mhz clock speed and multitasking capability. Each server is supported by a 170 MB magnetic disk for the temporary storage of image data files and database files.

4.2.3.1 Indexing

Creation of an image file index is the second major step in the Tennessee CMSR conversion. ODISS as presently configured has detailed indexing capability only for CMSR records.

At the index workstations, operators create a database entry for each file to allow search and retrieval based on name, regiment, company, and beginning and ending rank. First, middle, and last names of the soldier are divided into three separate alphabetic fields. Values for regiment, rank, and company fields are supplied by numeric code tables. There are 204 regiments in the Tennessee cavalry CMSR records, and up to three companies per regiment. The rank code table has thirteen different values. The index for each file also contains a remarks field. This field is displayed when the file is retrieved, but is not searchable. Since jackets were created at the regiment level and a single individual may have served in more than one company, an important use of the remarks field is to provide a cross reference to other files corresponding to service for the same individual.

Files are available for indexing as soon as file blocks are closed at the high-speed scanner station. Files arrive at index workstations in approximately first in/first out order. Indexing can ordinarily be accomplished by viewing the document image; it is rarely necessary to retrieve the paper record. After completing the index for the file, the operator enters the file's index record into the index database. This action automatically calls the next file to the workstation.

4.2.3.2 Quality Control

Quality control is the third major step in the conversion subsystem. As ODISS is presently configured, the quality control process is dedicated to CMSR files. In the Tennessee CMSR conversion, ODISS quality control had two purposes. Firstly, it was used to identify and immediately correct indexing mistakes such as misspelling of names or entry of the wrong numeric code for company, regiment, or rank. Secondly, it identified and marked in the

stored image file all images which needed to be rescanned and enhanced at the low speed scanner/rescan station.

CMSR records arrive at the quality control station in blocks of 40 to 60 CMSR files created by the high-speed scanner workstation. Blocks normally consist of all files placed in one Hollinger box during document preparation. At the quality control workstation, operators work with both paper and digital images of files. Images appear on the left side of the screen; index fields appear on the lower right side of the screen. After the operator retrieves a file, images in the file are displayed in the order they were scanned.

The jacket is the first item in each file. The operator selects the paper file which corresponds to the image file and compares index information on the jacket with index entries on the lower right side of the screen. To verify accuracy of values coded during the indexing process, code tables can be retrieved into a window on the upper right side of the screen. As operators are checking the accuracy of the indexing, they also evaluate the legibility of the jacket's image as the first item in the file. After finishing the index check, operators proceed through the file, comparing the paper to the digital image. If an image is illegible or of poor quality, the operator uses a function key to mark the image electronically for rescan. At the same time, the corresponding paper page is placed in a brightly colored folder, and the folder placed back into the CMSR file in the document's original location. If a page was missed during scanning, the document is also put into a colored folder and returned to its original place in the file. The proper page location is marked in the digital file electronically using the MISSING PAGE function key.

When quality review of the file is completed, the operator presses a function key to remove the file from the screen, build a table of poor images for rescanning and missing pages for scanning and insertion, and retrieve the next file in the block to the screen. When the last file in the block is completed, the operator returns to the initial quality control menu either to select another block or log off.

Unisys provided custom software for CMSR quality control which permits access to the Unify CMSR database for blocks and individual files. Quality control menus and function keys operate under Unisys programs, and blocks of files are accessed through the system manager software module.

4.2.3.3 Rescanning and Replacement

Scanning of previously unscanned images, including those missed by the high-speed scanner operator or those considered inappropriate for processing by the high-speed scanner, and rescanning of images of poor legibility is handled by a low-speed Ricoh platen-type scanner. This piece of equipment scans oversize documents measuring up to 11" by 14" at densities of 200, 300, and 400 dots per inch. It is particularly useful for capturing images from fragile documents, since the documents do not require physical movement through the scanner mechanism. After a document is placed on the stationary glass platen and the cover lid is put in place, the scan mechanism itself moves.

Images identified by quality control that require rescanning and fragile documents which were not processed by the high speed scanner are handled almost identically. File folders containing colored folders (containing the original documents from which the poor images were captured) are routed from the quality control workstations to the low-speed or rescan workstation. The operator calls up the appropriate file locating the image marked for rescan.

The operator then scans the document again, utilizing a greater variety of image enhancement techniques, until a better image is created. The new image replaces the poor one in the image file.

4.2.3.4 Retrieval

Requests for retrieval of digital images originate from workstations. These requests are sent to the system manager subsystem (see 4.2.5), along with the file control number corresponding to the images if available. The system manager retrieves the file control number from the index database if needed and routes the request to the initiation and monitoring subsystem. If image data resides in the archive subsystem, the request is passed over the multibus interconnect to the archives subsystem for servicing. Images are returned from optical disk storage to the archives subsystem server, which sends requested image data over the multibus either to a workstation for screen display or the print subsystem, where hard copies are furnished.

4.2.3.4.1 Staff Retrieval

ODISS includes two workstations for staff retrieval. These stations were intended for gathering data about the feasibility of having NARA staff perform CMSR searches using ODISS to reply to mailed-in inquiries for genealogical information.

To retrieve a file, staff members are first prompted to supply as many of the CMSR search fields as are known from the information in the mailed-in request: last name, first name, middle name, code values for rank in, rank out, regiment, and up to three companies. Fields are left blank if the information is not known, and the search is performed based on available information. A function key controls the beginning of the search. The system returns a scrollable list of matches (known as "hits"), or a message indicating that nothing was found. For each match, the file control number, complete index information, and number of images in the file are displayed on the screen. Viewing of images stored in the file is controlled by function keys which retrieve the file; rotate images filmed sideways as a result of size, or upside down as a result of original orientation on a two-sided document; zoom images; move directly to file image by number; and print either the hit list, all images in a particular file, or designated images within the file. When prints are made, the system calculates the cost of copies and produces a cover sheet listing the file control number, number of pages printed, and cost of the copies. Print options include a batch mode so that NARA staff can gather into one group a number of paid orders for copies and print all of them in a single operation.

4.2.3.4.2 Public Retrieval

ODISS includes one public workstation, designed for self-service reference of stored images. Workstation display screens are designed to guide the general public in the use of function keys and code tables to construct searches, retrieve files, and print index lists and file images. After reviewing these on-screen instructions, the public follows the same procedure used in the staff retrieval workstation, and has access to the same functions. When a public user decides to print hardcopies, the system notifies the user of the copy cost and allows the user to choose between stopping and continuing.

4.2.3.4.3 Remote Retrieval

An index-only, remote-site ODISS workstation was installed in the Tennessee State Library and Archives in Nashville. The workstation is linked to the ODISS system manager via 1200/2400 baud modems and a dial-up, voice-grade telephone line. The remote workstation includes a personal computer and dot matrix printer. Once data communication is established, the remote workstation has access to the system functions in the same manner as an on-site retrieval workstation. The remote system can query the index database, receive a hit list, and generate image print requests. Images are printed by the ODISS printer subsystem (at the National Archives), and the copies are mailed to the requester. Image data is not transferred to the remote workstation; researchers use the microfilm copy of the Tennessee CMSR available in Nashville to retrieve images based on information provided by the ODISS index hit list, and decide based on an inspection of the microfilm which ODISS CMSR images should be printed.

4.2.4 Archive Subsystem

The optical storage system consists of one optical disk autochanger with one internal drive controller, and two internal optical disk drives utilizing Sony 2.2 gigabyte, 12-inch optical disks. This system is daisy-chained to an external controller and two external drives. The external drives are used to write image data onto optical disk and to create backup security disks, so that the jukebox could be dedicated to retrieval of stored images. Both the drives and the controllers are themselves controlled over a small computer systems interface (SCSI). The SCSI bus carries all information to and from the writable disk controller and jukebox, and the SCSI interface includes all the commands necessary for complete control of these devices. Interface between the controller and external disk drives is accomplished by a proprietary Sony communications bus.

4.2.5 System Manager, and Initiate and Monitor Subsystem

4.2.5.1 System Manager Terminal

The system manager terminal is used to maintain and control data on employees, workstations, and index and image data stored on magnetic disk. From the System Manager Main Menu, eight basic database functions are available:

- Code table maintenance
- CMSR/non-CMSR file maintenance
- User Type maintenance
- Employee profile maintenance
- Workstation assignment maintenance
- Main reports menu
- Archives management
- Database backup read and write

Code table maintenance allows the user to enter or modify tables controlling codes for war, state, service, status, rank, regiment, and company. CMSR/non-CMSR files maintenance allows the user to add, delete, query or modify file indexes. User type maintenance controls the cost of prints. Employee profile maintenance allows the system manager to control individual access level to ODISS. Workstation assignment maintenance controls the workstation functions which a specific terminal or workstation can perform. The main reports menu controls the output of management reports automatically generated when ODISS is running. The Archive Management submenu is used to initiate the writing of a block of files to disk, display the status of the last file block written to optical disk, list blocks currently ready to archive, and find the total available space on the optical disk currently being written. Read and write database backup options provide backup and restore capabilities on the magnetic streamer tape used to backup the index disk database.

4.2.5.2 CSE/ARS Terminal

The CSE/ARS terminal is used to control the capture storage element and the archives storage functions of ODISS. Under the CSE terminal are offered eighteen functions which return information at the file and page (image) level on data stored on capture server element magnetic disks, initialize the disks, delete files, close files, display storage remaining on each disk, or perform other operating system functions. Under ARS, information pertinent to images already transferred to optical disk is made available. Six functions are available, including a directory function to read the contents of the directory of an optical disk.

4.2.5.3 IMS/Archive Control Terminal

The IMS/Archive Control terminal is used for two types of system-related actions. Its principal purpose is to initiate processes on the initiation and monitor subsystem. These include nine functions:

- Creation of a directory of file control numbers for each side of an optical disk. This is the directory which can be viewed under the ARS function. The directory is completed only after the entire side of the disk has been completed.
- Creation of a duplicate copy of an optical disk.
- Deletion of a previously archived file from optical disk.
- Dismounting of an optical disk from the jukebox or from one of the drives.
- Mounting of an optical disk.
- Retrieval of the index stored within each archival file.
- Initialization of an optical disk.
- Reading of a disk for errors.
- Check of volume status, i.e., what disk volume is loaded on a drive.

CHAPTER FIVE

ODISS TEST PLAN DESCRIPTION

5 ODISS TEST PLAN DESCRIPTION

This chapter presents the data collection and testing methodology used to gather ODISS operational statistics. Factors such as project test goals, data collection techniques, testing locations, and chronology are presented.

5.1 Testing Goals

A test plan was formulated to facilitate the collection, measurement, and evaluation of ODISS performance data. This plan established a structured testing process and provided management and staff with guidance in capturing test data and recording and analyzing results. ODISS supported analysis of the feasibility, costs, problems, and benefits of archival digital imaging systems. Workflow processes were evaluated, utilizing the system's inherent flexibility. System testing provided insight into quality control requirements for electronic imaging conversion projects, and public reaction and user acceptance to electronic images.

5.2 Test Sample Selection

Assorted NARA holdings were selected for testing, including Tennessee CMSR holdings, non-CMSR records, and microfilmed holdings. Tennessee CMSR records were selected because of their popularity with researchers, and the facts that the holdings had already been microfilmed and its size was suitable for a test environment.

Non-CMSR test document selection was based on a previous NARA document sampling and evaluation project. This sampling effort, conducted in 1985 during development of the *National Archives 20 Year Preservation Plan,* identified documents according to age, condition, and other image characteristics. ODISS project staff examined the test population outlined in the preservation study, and document samples were obtained for ODISS use. Ad hoc documents were selected according to physical construction, appearance, potential longevity, size, thickness, paper and ink colors, visual contrasts, and overall stability and need for repair. Documents previously microfilmed were useful in comparing digitally scanned imagery from documents and microforms.

5.3 Test Sample Attributes

Physical characteristics of the Tennessee CMSR (RG 109) varied with in size, format, texture, color of paper and ink, condition, and legibility. A survey conducted between 1982 and 1983 showed that the compiled military service records included cards in relatively good condition, and other documents with relatively minor problems. The majority of documents processed by ODISS were created by the War Department. Specifically, these were service jacket, reference slips, reference cards, and statement of service reference slips. There were also envelopes for specific documents, and folders for medical reference slips. Reference slips were printed forms, annotated with copyist handwriting in the early 1900's, and are generally of high ink-to-paper contrast.

5.3.1 CMSR Documents

The quality of original CMSR documents varied considerably, and the document paper quality was generally poor. The printed forms varied from off-white to dark brown, and the Confederate field operations and routine correspondence used blue-colored, ruled-lined paper.

The paper stock varied from coarse surface texture to very thin, tissue-like paper. Ink fading and bleed-through was common, and required special scanner processing.

Ink quality was not consistent, with only thick lines still visible on some documents. Endorsements were often fine line ink or pencil due to space restrictions, with occasional splattered ink blobs. Document sizes ranged from half-page up to letter and legal, with some larger sizes as well. Some documents consisted of several papers glued together. Double-sided, multiply folded pages were also common.

A volunteer soldier's compiled military service record was abstracted onto cards from muster and pay rolls, rank rolls, returns, hospital and prison records, and other military records. Access was through numerous card name indexes to the various series. A CMSR file typically consisted of combinations of the following:

- JACKETS: Jackets were heavy paper stock envelopes designed to hold the other documents. Jacket paper was often discolored, due to soiled and aged materials, and occasionally had reference information written in pencil on the flaps. Jackets contained the soldier's name, rank, etc., in fountain pen and ink technology which created fine thickness variations.

- REFERENCE SLIPS: reference slips were discolored paper stock which obscured the thin, fine handwritten lines, affecting the contrast ratio of ink to paper.

- REFERENCE CARDS: reference cards were typically heavily discolored. Reference cards were single cards which refer researchers to other places in the CMSR files for the actual documents. Reference cards were often created to account for the variations in the spelling of soldiers' last names.

- STATEMENT OF SERVICE REFERENCE SLIPS: statement of service reference slips are glued together at the top if they contain more than one page. The final page was usually a carbon of correspondence summary.

- OTHER ENVELOPES/FOLDERS: other envelopes for documents were often faded gray, due to the age and paper formulation used. The medical card folders were generally unfaded.

A second document category was Union and Confederate service-related records. The Union documents related to a soldier's status as a Prisoner Of War, including any parole time served. These were usually printed on good quality paper, although some are on very thin, fragile, tissue paper. The Confederate documents all relate to service in the Provisional Army of the Confederate States (P.A.C.S.). These documents ran the gamut of size, color, quality, and condition, occurring mainly in the files of officers, who typically were originators or receivers of provisional supplies:

- Printed forms, for requisitions (forage, clothes, equipment, etc.), pay accounts, discharge, etc.

- Hand-drawn forms for the above purposes.

- General Correspondence

The quality of these documents varied due to rough handling and poor storage under original field conditions and long-term retention. Many documents were stained and soiled, especially along fold lines. Due to brittleness, many documents were inserted into polyester sleeves during document preparation.

5.3.2 Non-CMSR Documents

In order to test the ODISS capabilities fully, the CMSR sample was supplemented with documents from other holdings. NARA's 20-year preservation survey was a valuable aid in document identification and selection. Documents exhibiting characteristics such as varied ink colors, faint images, brittle and varied colored papers, various fonts and typefaces, and others were identified. Production considerations such as image quality, document handling, and throughput rates were also identified and tested. High speed scanner conversion rates were evaluated with the various test documents. Government Printing Office holdings within NARA were also surveyed to identify suitable conversion candidates. Representative technical manuals were selected for high speed conversion and image quality evaluations. Other randomly selected documents, either from NARA holdings or provided by outside sources, were tested throughout the ODISS conversion effort.

5.3.3 Microform Samples

NARA has large microform holdings, and analysis of ODISS's ability to handle microforms was tested. The multiformat scanner accepted 16mm and 35mm roll films, 4 X 6 inch microfiche, and engineering aperture cards. All of these formats were tested to determine the equipment's ability to process NARA microforms. CMSR roll films were obtained and scanned, and quality comparisons were made between images captured from original paper documents and from film copies. Government Printing Office records which are currently undergoing microfiche conversion were also evaluated.

5.4 Testing Facilities and Locations

ODISS was tested prior to delivery to verify operational capabilities. The majority of ODISS data collection occurred in the ODISS laboratory facility in room B-31 of the Main Archives Building. Additional data were obtained from the other site locations for the public, staff, and the Nashville, Tennessee workstations. A public terminal was installed in NARA's Microfilm Reading Room (Room 400), while the staff terminal was installed in area 7E1. The public terminal was tested with the aid of walk-in users from the general public, while the Nashville remote site terminal was tested by Tennessee State Archives staff. All additional testing equipment and tools such as imaging test targets were obtained from various sources as required in support of specific tests.

5.5 Test Duration

The ODISS system was subjected to factory on-site testing prior to equipment shipment. This testing evaluated the system's ability to meet NARA requirements and validated the overall integration level. Testing held during the document conversion process evaluated operational factors and remote station access. Expanded Non-CMSR testing occurred following completion of the CMSR Cavalry records conversion activities. Test data acquisition began on September 2, 1988 following system acceptance testing and terminated on September 30, 1989 with the completion of the processing of Non-CMSR samples.

5.6 Constraints and Considerations

One ODISS test philosophy element was that the impact of component failure, or the performance of any one specific equipment item should not impact or influence any similar system components. Anomalies unique to ODISS hardware, software, or procedures were isolated when possible. Sample defects and unusual results were analyzed to determine if the cause was sample-specific, or were integrated design deficiencies.

Operational procedures were evaluated to increase understanding of optimum system configurations. The ability to reconfigure the existing system design allowed testing of alternative workflows. Factors such as scan density, image display resolution, and image enhancement algorithm requirements were analyzed to determine suitability for future archival applications.

During the testing sessions, system test conditions were monitored for compliance with ODISS test plan standards, test session data was recorded, unusual equipment or personnel conditions were noted, and usage of alternative software or hardware which would impede ODISS routine operations was avoided.

5.7 Measurement of User Satisfaction

Measuring user satisfaction with ODISS was accomplished with survey questionnaires, subjective assessments, and simulated database queries. User input centered around image quality, speed of data retrievals, and ease of system use for conducting information searches. The image quality analysis section of this test plan addressed image legibility, while the public/staff reference section presented hardware and software ease-of-use criteria.

5.8 Data Collection and Analysis Methodology

ODISS automatically collected considerable production data useful for monitoring routine operations. Other information was obtained using analytical testing, augmented with production staff and system user experiences.

5.8.1 Test Criteria Framework

Test criteria included test frequency, output formats, and procedural guidelines. Each subsection is identically formatted with a factor, method, procedures, and test sequence. These criteria are described as follows:

- FACTOR: Presents the criterion to be tested.
- JUSTIFICATION: Provides the reason(s) for including the factor in the test plan.
- METHOD: Brief description of the planned test approach.
- PROCEDURES: Provides methodology to be followed:
    - Test sequence: Procedural guidelines to be followed to conduct that particular test criterion.
    - Test frequency: Planned frequency of testing for that particular criterion.

- ¤ Output format: Medium in which testing data will be provided or accessed.
- ¤ Data analysis: Methodology to be used in analyzing compiled test data.
- ¤ Supplemental: Any additional or correlated pertinent analytical information.

✲ COMMENTS: Any additional information considered useful for the testing process.

The above criteria are included as needed in the following test plan procedures.

5.8.2 Test Criteria Descriptions

The following test plan criteria were utilized during the testing and data collection phases of the ODISS project.

5.8.2.1 High Speed Scanning

A. FACTOR: Production rates for number of images scanned and files processed.

JUSTIFICATION: Production rate information is important for estimating future equipment requirements.

METHOD: Automatic management reporting for day, week, month, quarter, and year data.

PROCEDURES:

Test sequence: Obtain scanner production statistics from system manager, analyze data and draw conclusions, observe production techniques.

Test frequency: Continuous and automatic.

Output format: System manager terminal displays and printed reports.

Data analysis: Automated statistical analysis.

Supplemental: Observation of station operation to identify staff or equipment deficiencies.

B. FACTOR: Fragile and/or oversize document processing with the high speed scanner.

JUSTIFICATION: This area supplements general production throughput rate data.

METHOD: Evaluate transport operations, and the scanner's ability to accept unusual document sizes.

PROCEDURES:

Test sequence: Critique scanner operations during CMSR conversion, supplemented with observations during ad hoc testing using various document types.

Test frequency: *Ad hoc* testing and observation.

Output format: Observer's recorded notes.

Data analysis: Subjective review of scanner transport operations.

C. FACTOR: Scanner equipment reliability statistics.

JUSTIFICATION: Information useful for equipment on-site service requirements.

METHOD: Analyze maintenance technician logbooks.

PROCEDURES:

Test sequence: Review hardware repair logs, personal notes, and discuss with the operations staff.

Test frequency: On-going observation of equipment operation.

Output format: Observer's notes.

Data analysis: Manual review of multi-source data.

Supplemental: Hands-on testing of hardware to evaluate adequacy of built-in status indicators and operator controls.

D. FACTOR: Scanner design to include: ease of use by a single operator, display monitors and keyboards, document catcher operation, and pushbutton controls.

JUSTIFICATION: Useful in estimating future staffing and need for special system features.

METHOD: Scanner design review and operations analysis.

PROCEDURES:

Test sequence: Note problems encountered with attention directed to oversized document handling, display monitor usage, access to document catcher bin, and operator's control panel.

Test frequency: *Ad hoc* test sessions and periodic summarization of experiences.

Output format: Observer's comment sheets.

Data analysis: Compare one and two-person operations.

5.8.2.2 Image Quality

A. FACTOR: Imaging analysis to include the effects of: documents in polyester sleeves; various ink colors and document paper qualities (stains, bleed through, dirtiness).

JUSTIFICATION: Data useful for estimating future scanner image processing requirements.

METHOD: Scan in samples and analyze screen image and print qualities.

PROCEDURES:

Test sequence: Calibrate scanner, scan test targets and analyze system performance. Record results and any unusual test conditions.

Test frequency: *Ad hoc* testing as required.

Output format: Workstation screens and laser prints.

Data analysis: Subjective comparison of image qualities under various testing conditions.

Supplemental: Use of both internal and commercially available test targets.

B. FACTOR: Addition of optical lens filters for improved image quality.

JUSTIFICATION: Scanner's ability to capture all ink colors is important.

METHOD: Analyze impact of filters on image quality.

PROCEDURES:

Test sequence: Visually compare images scanned with and without various lens filters. Determine best filter combinations for various record attributes.

Test frequency: *Ad hoc* tests.

Output format: Test notes.

Data analysis: Subjectively compare image qualities with different lens filters installed.

C. FACTOR: Scan density (DPI) needed to meet NARA needs.

JUSTIFICATION: Impact of potential storage savings using minimal scanning rate is significant.

METHOD: Comparative analysis of scan densities and image legibility.

PROCEDURES:

Test sequence: Scan document test batch on both high (200 dpi) and low speed scanners (200-400 dpi); utilize targets and NARA documents. Examine screen images for legibility. Print images for laser printer evaluations.

Test frequency: *Ad hoc* testing sessions.

Output format: Display screens and laser prints.

Data analysis: Subjective comparison of image qualities.

Supplemental: Scan specialized test targets.

D. FACTOR: Relationship of scanner contrast settings to image quality and digital image file sizes.

JUSTIFICATION: Station productivity and image quality are affected by equipment operations.

METHOD: Scan documents at various settings to identify optimum contrast settings.

PROCEDURES:

Test sequence: Capture documents using different automatic thresholding and operator controlled modes. Compare image quality and digital file sizes.

Test frequency: *Ad hoc* testing sessions.

Output format: Display screens and laser prints.

Data analysis: Subjective comparison of image quality and file sizes captured using automatic thresholding versus manual intervention.

Supplemental: Analysis of image storage sizes based on threshold setting.

E. FACTOR: Image quality comparisons of digital screen images and hardcopy prints.

JUSTIFICATION: Image legibility comparisons are useful system performance indicators.

METHOD: Structured testing sessions with NARA staff and professional researchers.

PROCEDURES:

Test sequence: Assemble group for "blind test" evaluations. Conduct tests using the following: original documents, NARA microforms, and various scanned images. Elicit responses relative to legibility, information completeness, and usefulness.

Test frequency: Testing and analysis as required.

Output format: Video screen images and hardcopy output.

Data analysis: Subjective comparison of image quality and legibility.

Supplemental: Introduction to the system will precede testing.

F. FACTOR: Image quality comparisons of digital images captured from paper documents to digital images captured from microforms.

JUSTIFICATION: Comparison of paper and film input scanning technologies is important for future decisions.

METHOD: Scan images from paper records and compare to images scanned from microforms.

PROCEDURES:

Test sequence: Obtain paper records and matching microforms. Scan the documents and the microforms; compare screen images and hardcopy prints.

Test frequency: *Ad hoc* testing.

Output format: ODISS workstation display screens and laser prints.

Data analysis: Subjective evaluation of image quality using paper and microform input.

5.8.2.3 Production Workflow

A. FACTOR: High speed scanner performance measurements

JUSTIFICATION: Scanner production rates are significant elements for NARA record conversions.

METHOD: Summarize data collected from timer programs.

PROCEDURES:

Test sequence: Collect production statistics using timer programs. Process data using timer software to determine work and wait times. Determine impact of system load on response times.

Test frequency: Timings conducted during ODISS operations.

Output format: Data files in timer software format.

Data analysis: Statistical evaluation of work time versus wait time measurements.

Supplemental: Analyze impact of file open and close operations on productivity.

B. FACTOR: Elapsed times for indexing and quality control operations, including impact of file retrieval on station performance.

JUSTIFICATION: Useful for distinguishing source of throughput problems.

METHOD: Analyze data collected using timer programs; work time versus wait time.

PROCEDURES:

Test sequence: Collect work time production statistics using timer programs. Process data using timer software to determine work and wait times. Determine impact of file open and close operations on productivity.

Test frequency: Timings conducted during routine ODISS operations.

Output format: Data files in timer software format.

Data analysis: Statistical evaluation of work time versus wait time measurements.

Supplemental: Analyze impact of system load on response times.

5.8.2.4 Indexing

A. FACTOR: Number of indexing files processed.

JUSTIFICATION: This is a basic ODISS productivity measurement.

METHOD: Automatic data collection.

PROCEDURES:

Test sequence: Obtain ODISS system management reports and analyze production data.

Test Frequency: Continuous data collection.

Output format: Screen display and system printouts.

Data analysis: Compare learning curves to on-going production.

Supplemental: Staff interviews and observations.

B. FACTOR:	Number and description of data entry/indexing errors.
JUSTIFICATION:	Useful for determining impact of indexing errors and possible system design changes.
METHOD:	Summarize quality control station experiences.
PROCEDURES:	
Test sequence:	Collect information concerning index error rates and causes.
Test frequency:	On-going observations.
Output format:	Staff questionnaires.
Data analysis:	Evaluate start-up versus on-going operations.
Supplemental:	Staff interviews.
C. FACTOR:	Ease/difficulty of code tables scrolling system.
JUSTIFICATION:	Useful for evaluating system design which requires key entry of numeric codes and line-by-line scrolling of code tables by operators.
METHOD:	Operator interviews, hands-on analyst use.
PROCEDURES:	
Test sequence:	Collect and analyze operations data. Operators scroll numeric code tables for fields such as rank and regiment. Key-enter numeric data; determine ease of use and operator's ability to learn system operations.
Test frequency:	*Ad hoc* testing as needed.
Output format:	Recorded notes and questionnaires.
Data analysis:	Subjective evaluation of comments.
D. FACTOR:	Personnel rotational assignments and impact on system operations.
JUSTIFICATION:	Varied work tasks can reduce tedium, and cross training provides more team skills and backup capability.
METHOD:	Review system operation under static operator and rotational staffing plans.

PROCEDURES:

Test sequence: Analyze system productivity for periods of routine operator assignments, compare with production statistics during rotational activities.

Test frequency: As required.

Output format: Notebooks and staff survey logs.

Data analysis: Observation and subjective comparison of static versus rotating personnel assignments.

Supplemental: Review of employee longevity and impact of people substitutions.

5.8.2.5 Quality Control

A. FACTOR: Station production rate for number of files and images completed at Quality Control.

JUSTIFICATION: Production data useful for future system design considerations.

METHOD: Automatic data collection by system.

PROCEDURES:

Test sequence: Obtain ODISS management reports for work periods required; analyze data.

Test frequency: Continuous.

Output format: CRT screen and printouts.

Data analysis: Statistical comparison of performance at start-up (with its learning curve) to performance during later, ongoing production.

Supplemental: Observations on ease of use.

B. FACTOR: Quantity of images rejected, and number of electronic place holder images created for documents not scanned.

JUSTIFICATION: This quantifies the error rate for pages missed.

METHOD: Automatic data collection by system.

PROCEDURES:

Test sequence: Obtain management reports and analyze data. Summarize the data collected by the system.

Test frequency: Continuous.

Output format: Display screen and printout.

Data analysis: Statistical comparison of start-up learning curve to ongoing operations.

Supplemental: Observations on ease of system use.

C. FACTOR: Special station ease of use features, such as image rotate, zoom, menus, function keys, etc.

JUSTIFICATION: Useful in design changes for improved workstation efficiency.

METHOD: Hands-on operations to acquire needed analytical data; interviews with experienced operators.

PROCEDURES:

Test sequence: Interview operators and analyze staff responses.

Test frequency: As required.

Output format: Analysts' notes and interviews with operators.

Data analysis: Evaluate data in terms of learning curve.

5.8.2.6 Low Speed Scanning and Enhancement

A. FACTOR: Low speed station production rates.

JUSTIFICATION: System productivity measurements require reliable statistics.

METHOD: Automatic collection of data for day/week.

PROCEDURES:

Test sequence: Print out daily/weekly report data for low speed station; derive quarterly and annual data; analyze results.

Test frequency: Periodic collection and printing of report data.

Output format: System manager screens and printouts.

Data analysis: Statistical evaluation of low speed productivity.

Supplemental: Evaluation of system load impact on station performance.

B. FACTOR: Scanner settings invoked for typical problems, and ease of use of image processing subsystem.

JUSTIFICATION: Identifying problem documents and image processing algorithms is important.

METHOD: Summary of operators experiences with image enhancement system.

PROCEDURES:

Test sequence: Compile operations experience concerning rescan operations; analyze image processing capabilities and ease of use.

Test frequency: *Ad hoc* testing.

Output format: Observer's and operator's experience logs.

Data Analysis: Evaluation of the low speed station.

5.8.2.7 System Manager

A. FACTOR: System's ability to collect, compile, and generate accurate management data and required reports under operational conditions.

JUSTIFICATION: Reliable system management reports are important to monitoring performance.

METHOD: Review findings and assess value of management reports.

PROCEDURES:

Test sequence: Monitor system manager report production schedule; verify reports for accuracy; assess utility of available information.

Test frequency: On-going report evaluations.

Output format: Display screens and printouts

Data analysis: Subjective evaluation of system's data collection and reporting capability.

Supplemental: Need for expanded reporting capability.

B. FACTOR: System manager's workstation design and ease-of-use factors.

JUSTIFICATION: The system manager is the central control point for the system and should be operated efficiently.

METHOD: Analyze user interfaces and station layout.

PROCEDURES:

Test sequence: Evaluate system manager workstation design, and its ability to perform required functions.

Test frequency: *Ad hoc* testing as required.

Output format: ODISS display screens and printers.

Data analysis: Observation and subjective evaluation of the system manager station design layout and ergonomics.

Supplemental: Weigh alternative system configurations for performing routine operations.

5.8.2.8 System Operations

A. FACTOR: ODISS's ability to perform non-CMSR item processing.

JUSTIFICATION: In order to meet project goals, it is important for ODISS to accept Non-CMSR documents and microforms.

METHOD: Process Non-CMSR documents and microforms.

PROCEDURES:

Test sequence: Use test batches on high and low speed scanners, and verify system's capability to accept, store, and retrieve the files.

Test frequency: *Ad hoc* tests.

Output format: System screens and printouts.

Data analysis: Subjective comparison of CMSR and non-CMSR item processing.

Supplemental: Evaluate search procedures for both item processing schemes.

B. FACTOR: ODISS system workflow design analysis.

JUSTIFICATION: Efficient and productive system operation depends on a balanced workflow process.

METHOD: Analyze alternative methods and hardware configurations.

PROCEDURES:

Test sequence: Experiment with and analyze alternative designs and production methods. Determine equipment and configurational needs for efficient operations.

Test frequency: As needed.

Output format: Screen and hardcopy prints.

Data analysis: Analyze alternatives to existing production workflows.

5.8.2.9 Microform Scanning

A. FACTOR: Verify scanner's image processing to handle typical quality microforms.

JUSTIFICATION: Quality digitized microform images require image processing capabilities.

METHOD: Scan microforms and analyze the results of image processing tests.

PROCEDURES:

Test sequence: Scan images under various scanner settings and compare results.

Test frequency: Special test series, plus *ad hoc* testing as needed.

Output format: Film scanner screen and laser prints.

Data analysis: Subjective comparison of captured images before and after enhancements.

B. FACTOR: Scanner ease-of-use to include controls, film handling, monitor(s) placement, keyboards, etc.

JUSTIFICATION: Film scanning productivity is important for any system requiring scanning from microform holdings.

METHOD: Analyze hardware and station operation; note unusual techniques.

PROCEDURES:

Test sequence: Observe film scanner human interface design during routine operations and testing sessions. Study the existing layout, and note ease-of-use features in the hardware or software interfaces.

Test frequency: Continuous.

Output format: Notes and observer's impressions.

Data analysis: Critical analysis of station ergonomic design and ease of use.

5.8.2.10 Index Storage

A. FACTOR: Index data storage requirements.

JUSTIFICATION: This information is useful in determining system storage requirements.

METHOD: Analysis of disk file data capacities and usage of system reports.

PROCEDURES:

Test sequence: Collect and analyze data concerning magnetic storage and index overhead.

Test frequency: Periodic.

Output format: Display screen and printouts.

Data analysis: Tabulation of total storage required to hold CMSR index records.

Supplemental: Determine magnetic storage remaining after conversion of CMSR records.

5.8.2.11 Image Storage

A. FACTOR: Image capacity of Sony CAV 12" optical disk.

JUSTIFICATION: Validation of image storage requirements capacity is important.

METHOD: Analyze disk image capacities and usage levels using system reports.

PROCEDURES:

Test sequence: Obtain and analyze system reports regarding ODDD space usage and image counts.

Test frequency: Periodic.

Output format: System printouts.

Data analysis: Review of system printouts for optical disk usage statistics.

Supplemental: Determine capacity of disks for larger-sized images.

B. FACTOR: Jukebox performance and the system's ability to service concurrent requests for image retrieval.

JUSTIFICATION: Jukebox performance is important for image retrieval productivity.

METHOD: Simultaneously request image data from several workstations; analyze jukebox server operations under loaded conditions.

PROCEDURES:

Test sequence: Simultaneously request image files stored on different optical disks from several terminals and observe system response. Record any problems the system has with file retrieval from the jukebox or with ensuing file transfer operations.

Test frequency: *Ad hoc* data collection.

Output format: Display screens and printouts.

Data analysis: Statistical analysis of jukebox's performance in servicing of user requests.

5.8.2.12 On-Site Reference

A. FACTOR: Workstation ease-of-use to include: keyboard features, functions, and terminal display.

JUSTIFICATION: Useful for improving user access.

METHOD: Questionnaire/interviews with system users.

PROCEDURES:

Test sequence: Conduct user interviews, training sessions; use questionnaires to gain information on station ease-of-use.

Test frequency: Ongoing data collection.

Output format: Observer's recorded notes.

Data analysis: Subjective evaluation of workstation overall design and human interface features.

Supplemental: Study the ease-of-use factors and their impact on productivity.

B. FACTOR: Software ease-of-use to include: menus and code tables, retrieving a file's images, returning to the index list, retrieving another file's images, etc.

JUSTIFICATION: Useful for improving ODISS and for the design of a larger production system to improve ease of use.

METHOD: Questionnaire/interviews with users.

PROCEDURES:

Test sequence: Work with randomly selected users to determine the software's ease-of-use and their comprehension of its retrieval functions and capabilities. Complete questionnaires and analyze results.

Test frequency: Ongoing data collection.

Output format: Questionnaires and observer's recorded notes.

Data analysis: Subjective evaluation of software design and station operations.

5.8.2.13 Remote Reference

A. FACTOR: Remote station menu design.

JUSTIFICATION: Information about the system's user interface is important for future system design.

METHOD: Data collection using interviews and/or written user evaluations.

PROCEDURES:

Test sequence: Collect data on ease of use and user-friendliness of the station's user menus. Analyze results.

Test frequency: Ongoing data collection.

Output format: Observer's recorded notes

Data analysis: Subjective analysis of software design and station operations.

B. FACTOR: Value of retrieving index data only; Tennessee State Archives user interest in receiving image data.

JUSTIFICATION: Costs of digital image transmission to remote sites should be weighed in comparison to user needs.

METHOD: Telephone interviews with remote site manager.

PROCEDURES:

Test sequence: Conduct interviews with remote users to gain needed information and compare to on-site operations.

Test frequency: As needed to gather data.

Output format: Interviewers' recorded notes.

Data analysis: Comparison of index data searches and manual retrieval of microfilm from Tennessee's local holdings of microfilm copies with ODISS's ability to retrieve both index information and document images from the same workstation.

C. FACTOR: Identification of system access problems (response time, sign-on timeliness, etc.)

JUSTIFICATION: Data concerning remote users access is important in deciding scope of access in any future, expanded system.

METHOD: Log recorded by remote site users.

PROCEDURES:

Test sequence: Conduct periodic phone interviews with Tennessee State Archives staff to gather information about access experiences and problems.

Test frequency: As needed to gather information.

Output format: Recorded notes.

Data analysis: Analyze system contention problems.

5.8.2.14 Hardcopy Output

A. FACTOR: ODISS system laser prints: overall quality and legibility.

JUSTIFICATION: System output quality must be legible, even for small type point sizes.

METHOD: Print and analyze samples on system laser printers.

PROCEDURES:

Test sequence: Print and examine hard copies from a number of documents with a wide range of document characteristics. For critical comparisons, use files captured from standard test targets.

Test frequency: Periodic, in conjunction with various types of documents processed.

Output format: Laser print output.

Data analysis: Subjective comparison of image prints based upon image quality criteria.

B. FACTOR: Hardcopy laser prints compared to screen images.

JUSTIFICATION: Hardcopy replications must be as good or better than those rendered on high resolution screens.

METHOD: Comparison of sample test target prints to high resolution screen images.

PROCEDURES:

Test sequence: Print previously scanned files on ODISS laser printers; compare printed copies to screen images; analyze differences and any significance.

Test frequency: *Ad hoc* testing and observations.

Output format: Screen displays and laser prints.

Data analysis: Visual comparison of 150 dpi screens and 400 dpi laser prints.

Supplemental: Make use of test targets which contain special features useful in analysis.

CHAPTER SIX

PROJECT OPERATIONS ANALYSIS
AND
TEST RESULTS

6 PROJECT OPERATIONS ANALYSIS AND TEST RESULTS

The purpose of the ODISS project was to gather information concerning feasibility of using digital imaging systems in support of archival programs and operations at the National Archives. In order to answer that question, several approaches were implemented including monitoring routine operations during typical production, and system testing under controlled conditions. The following subsections describe the experiences and knowledge gathered during the ODISS system operations and performance testing. The data collected during performance and operational investigations are also discussed. Test data and ongoing operational results analysis are provided for each major subsystem. A discussion of the important issue of image quality and the results of intensive public and staff image analysis sessions is also included.

It should be noted that because of the unique design of the ODISS system, actual throughput performance achieved with the ODISS system which was designed in 1986, does not fully reflect the capabilities available with newer digital imaging technology as currently marketed. For complete descriptions of the ODISS system hardware, software, and operating procedures, refer to Appendix B.

6.1 Document Preparation For The ODISS Project

Before paper records are microfilmed for National Archives publications they are prepared for filming. Similar document preparation work was performed on the Tennessee Confederate CMSR records to get the files ready for digital conversion through the ODISS input processes. While the document preparation of the Tennessee CMSR records was essentially the same as the traditional work done for microfilming, there were some features unique to ODISS.

6.1.1 Tennessee CMSR Records

Document preparation involves putting the records in order for conversion. This includes flattening folded papers, removing such fasteners as staples and paper clips, correcting any misfilings so that the documents are arranged in the proper order, and making any necessary new box and folder labels. Any special preservation problems are identified, and in general proper preservation procedures are followed as outlined in the Archives guidelines for holdings maintenance.[61] Production standards for individual workers are set for each document preparation or holdings maintenance project. Work is done in batches, which are checked by supervisors for quality and statistical errors. The size of batches and the time required to complete a given amount of work fluctuates widely between projects because the standards for each project depend on the characteristics of the records and any other special features unique to that particular project. So, for each project time and production standards are set and written instructions are generally prepared.

These normal procedures were followed during the document preparation of the Tennessee Confederate CMSR records for ODISS. Written instructions for the project were developed

[61] Mary Lynn Ritzenthaler, *Preservation of Archival Records: Holdings Maintenance at the National Archives* (1988).

to guide the work.[62] A production standard of completing 7.5 to 7.6 boxes per day was set. An existing Tennessee CMSR records storage box contained approximately 150 to 180 soldiers' files. After document preparation, the contents of each "old" box required three "new" boxes because of the addition of file folders and the flattening of documents.[63]

During document preparation, an error rate of six or more was enough to fail the batch. Batches were defined as three of the old boxes containing the CMSR files. Errors were defined to include such housekeeping details as failure to fill out the time or other date correctly on the batch sheets and such substantive matters as improper labels, improper placement of documents in folders, and failing to put the indicators for two-sided scanning on the appropriate documents.

The Tennessee CMSR records were kept in the current order by regiment with the files for all members therein arranged alphabetically by surname. The major work involved the preparation of the documents in each file. Each file had its own new folder. The documents were removed from their envelopes or jackets, and the jackets were placed first in each folder with their flaps opened. Next the standard size CMSR regimental cards were placed after the jacket. Then any other documents that might be in a file were put in the folder and flattened if necessary. So, the order of the documents in the folder was flattened jacket first, then regimental cards, and finally flattened loose documents.

Many of these other loose documents had been tri-folded to fit into the jackets and needed to be flattened. This was a significant factor in the great expansion in the space needed by the series; after document preparation the Tennessee CMSR records occupied about three new boxes for each old box. Each new box usually held fifty to sixty files although the varying number of documents per file meant that boxes might have fewer than forty or, in very rare instances, as many as ninety folders. To mark documents for two-sided scanning, plastic clips were put on jackets and cards were placed in the folders with the reverse side facing forward. Whenever possible, the documents after the cards were arranged by size from smaller to larger. Fragile documents were placed in polyester sleeves. After the records were reboxed, new permanent labels were prepared using a computer program running on a laptop computer.

The Tennessee document preparation had a larger staff than is usually available for such projects. Many of the people hired to run the ODISS equipment began work well before the delivery of the system and were assigned to preparing the Tennessee records. The preparation staff assigned to the Tennessee Confederate Compiled Military Service Records was composed of one archivist, who oversaw the project and served as liaison to management; one senior archives technician, who acted as quality control reviewer; and a staff of archives technicians, varying in number from two to thirteen, who performed the bulk of the actual document preparation. The staff included, at times, two detailees from the Records Declassification Division (NND) and as many as six persons hired to work in the ODISS laboratory.

[62] See "Instructions for Processing Tennessee Confederate Compiled Military Service Records for the ODISS Project."

[63] During the scanning of the Tennessee Cavalry CMSR records, the soldier's files were found to average four images each.

6.1.2 Differences Between Preparation for ODISS and Microfilming

Because the document preparation workload is a significant factor in starting a project, it was important to analyze similarities and differences between document preparation requirements for digital imaging and microform conversions. NARA has extensive experience with preparing documents for microfilming, and this experience was compared to the preparation of the Tennessee CMSR records for the ODISS project.

Since only the Tennessee CMSR files from Record Group 109, the War Department Collection of Confederate Records, were prepared for digital scanning, it is difficult to say exactly whether the differences apply to other groups of records. It may be more practical, however, to categorize the highlights noted. For example, microform document preparation does not require arrangement of documents by size or color, nor the electrostatic copying of original documents which would be useful for ODISS. Some of these requirements for arrangement would help maximize scanner throughput in ODISS whose scanners had manual exposure and size adjustments, functions which may be automated in future scanners.

ODISS document preparation, on the other hand, did not require drafting title pages, introductions, tables of contents, targets, special lists, roll notes, and appendices that are produced for microfilm publications. The beginning and end of film roll targets which are placed in document series destined for microfilming, also have no ODISS equivalent. Nor did ODISS require the procedure for microfiche production, where the pages are usually counted prior to filming so that fiche breaks and numbering for titling can be performed.

Many of the above filming requirements are due to the lack of automated indices in the typical NARA microfilming publication. The ODISS computerized index operation facilitated the data capture and subsequent image retrievals in an automatic mode, rather than relying on manual look-up techniques required for non-automated filming systems.

6.1.3 Lessons Learned

The preparation of the Tennessee Confederate CMSR records for ODISS demonstrated that the normal basic procedures used in projects for microfilming and other holdings maintenance efforts work also in preparing documents for digital scanning. Written guidelines, time and production standards, and most of the same document handling operations employed in microfilming projects were applicable to document preparation for digital scanning.

There also were some differences. As described in the previous section many of the steps necessary for microfilming are eliminated in the document preparation for digital scanning whose final product includes an automated index. Another area of difference is the grouping of documents by size or color to reduce the need for making manual adjustments at the scanner. While this was not done much in the ODISS project, it might be more feasible with other bodies of records and it could be useful for any scanner that is dependent on manual adjustments for variations in thresholding and other image capture techniques.

In summary, the ODISS experience showed that while document preparation for digital scanning may have some differences with more traditional document preparation, the Archives' wealth of experience in this area is a sound foundation for preparing older records for the new digital technology.

6.2 High Speed Scanning

6.2.1 Ease of Use of the Workstation

The high speed scanning workstation consisted of the TDC/Photomatrix scanner, the file control terminal, the high resolution controller terminal, and the supplemental scanner control terminal. Since throughput speed was of paramount concern for this station, most aspects of the determination of the ease of use of the workstation fall into this category. In any production type environment, redundancy of operator actions will speed up the process. Also, fewer operator decisions at this stage tended to speed up throughput of documents through the scanner.

The scanner itself proved to be quite easy to use, once the operator was trained on the functionality of the controls and given a day's worth of supervised practice. The control panel at the front of the scanner was useful for changing size masking and controlling contrast levels. The three most common sizes could be identified with the touch of one button. Other sizes were activated by use of the supplemental control terminal that sat on top of the scanner. This required four keystrokes for each size.

Contrast control could be separately maintained for both the top and bottom scanner arrays. For instance, if a two-sided document had a dark side requiring light contrast and a light side requiring dark contrast, contrast settings could be made for each scanner array (top and bottom) in order to optimize image quality.

Standard operational file management control could also be easily maintained from the scanner control panel. Blocks and files within blocks of work could be opened and closed with the touch of a single button. ODISS used a button on the scanner control panel to indicate the need for the system to open a block and, thereafter, a file. This setup was quite easy for the operator to use in a production mode. When the first page of a new file was ready to be scanned, the operator could close the previous file and open a new file simply by pushing the OPEN FILE button. The operator could easily indicate either the need to scan one or both sides of the document.

The control panel also had a small status screen that would indicate the status of the scanner. On a normal input sequence, it would show the image count. If a jam or some other problem occurred, the status would be displayed instantly so that the operator would not need to spend unwarranted time trying to discover what had caused the scanner to stop.

The file control terminal supplemented the scanner control panel for file and block control. The block and file numbers were prominently displayed with three-inch numerals. They were displayed as either crosshatched or solid to indicate the file open status. The terminal also provided additional controls for restart, reset, and other situations that could arise during a scanning operation. It is important to have this kind of control and involvement for input systems requiring file maintenance.

6.2.2 Production Rate and Throughput

The production rate of the high speed scanning operation is a function of the combination of many factors. For the Tennessee CMSR test sample, ODISS's high speed scanner was able to capture as many as 3300 images per hour. While speed of the scanner itself was the

primary item for analysis, many other factors that affected the average rate of speed were important to consider.

6.2.2.1 CMSR Sample

There were certain common characteristics of the CMSR sample that affected the high speed scanning operation. Some were beneficial and others detrimental. There were four main categories of document types within the CMSR sample, the CMSR jacket, the cross-reference card, the file card, and supporting pages. The first three categories are generally standardized, within their own group, in size, color, and condition. The supporting pages are made up of all other types of documentation that would commonly comprise a personnel file. Requisition forms, pay vouchers, and letters constitute a large portion of this category. These have little commonality and required the operator to adjust the size masking and threshold control of the scanner for virtually every page.

The jackets and cards had pre-set size settings that were invoked by a single button. By using the masking feature, the image was cropped to a specific, pre-determined size in order to limit extraneous image capture resulting in inordinately large image file sizes. In the case of the jackets and cards, they were smaller than an 8.5" X 11" size which would normally create a large black border around the remaining areas. With size masking, not only was the file size reduced, but the large black borders were eliminated, saving toner on prints.

In the situation when oversized items or any item with an unscannable condition came up in the file, a substitute page was scanned in its place and it was refiled with the original file pages in the hopper. When the "place holder" was encountered during quality control, it was marked for rescan.

6.2.2.2 File Control Considerations

There are two basic methodologies that can be followed when designing a digital image-based document retrieval system. The first allows for access to the lowest level, the individual image, directly without going through any hierarchical levels. The second permits access to the image level, but only after the file has been accessed. The analogy could be made of the difference between having a file drawer full of individual pages and picking a single page to view as opposed to the same file drawer having, this time, the pages in a number of folders. In order to have a non-random access to the page, the file must be selected first. At that point, the page may be selected directly.

Most digital image-based, record management systems utilize the file access instead of page access methodology. If the file level is to be accessed, the beginning and ending of the file must be identified. In ODISS, individual page images are directly accessible, but only after the file has been retrieved. Sequential page numbers were automatically attached to each image within a file during the scanning input. This allows the user to access the file and then jump directly to a particular image number within that file. This capability proved to be an excellent compromise between only file access and direct page image access. Preparation for direct page image access would be very costly in terms of time, since each image would have to be indexed individually.

The entire ODISS conversion subsystem was controlled by the opening and closing of blocks and files for each operation. This enabled the system manager subsystem to eliminate the chance of duplicate and simultaneous actions to be taken on the same file. It also served to

maintain and force the correct sequencing of documents through the input conversion subsystem. In other words, two terminals should not be indexing the same file. Or, a file must have been indexed before it goes to quality control. Without file and block control, it would be almost impossible to control the movement of work through the production line.

The detrimental aspect of opening and closing files and blocks at each operational station throughout the conversion subsystem is one of time. It takes time to create files and blocks, to allocate them disk space, and to open and close them on an irregular basis.

Unisys designed the system throughput to conform with the estimated file sizes as given in the Invitation For Bid (IFB). This estimate was that the average file size for the Compiled Military Service Records (CMSR) was fifteen images. This meant that, for the high speed scanning component, the operator would be able to scan approximately fifteen images into the file at a high rate of speed before having to open a new file. It turned out, however, that the actual average number of images per file was four. Therefore, operators had to wait on file openings and closings almost four times that which had been anticipated. Larger file sizes would have enabled the operator to scan documents at a rate much closer to the projected speed of 40 images per minute since the wait time for file openings and closings would have been reduced.

6.2.2.3 Handling Image Anomalies

Standard office-type documents offered no real challenge to the ODISS high speed scanner operation. The legibility and general quality of the images were excellent and documents could generally be scanned without frequent contrast level changes. Jackets and cards from the CMSR sample could be scanned with standard and fairly constant settings at high rates of speed. There were, however, some document characteristics not easily handled by using particular consistent settings. These documents required changes in contrast levels in order to attain the best possible image quality. In some cases, the document was too deteriorated for the standard constant thresholding-type of image processing to be able to produce adequately acceptable image quality.

To be efficient during a high speed input scanning process, the documents must be scanned with a large degree of procedural repetitiveness. That is, the input operator should not stop the production sequence in order to walk around the scanner, retrieve the document, and rescan it using different settings. The ODISS method of handling this situation was to scan everything through the high speed scanner using the best predetermined setting possible, as gained through experience. Each image was displayed on the high resolution screen mounted above the scanner as soon as it was scanned. The operator was to monitor the quality of the image only at a rate of every five to six images in order to maintain concentration and paper feeding speed. Only five percent of the images created on the high speed scanner were rejected in quality control due to image quality deficiencies. By letting the relatively few poor images through the high speed station, the throughput was maintained. In those cases where the image was rejected in the quality control operation, the images were rescanned using a tabletop scanner in an off-line sequence where the level of throughput was not nearly as high. In this situation, more finesse could be used with image enhancement techniques in order to achieve the best possible image quality. For almost 95 percent of the sample documents, the extra processing was not necessary; and as a result, the operational approach taken with the high speed scanning procedures was successful.

6.2.2.4 Pension and Bounty Land Warrant Sample

The Pension and Bounty Land records along with the Compiled Military Service Records comprise 34 series of files totalling approximately 335 million images and over 10 million individual files. These records are relatively active as they collectively are highly referenced by genealogists. Since the main body of records used in the ODISS tests came from the CMSR series, an NN-supported holdings survey was used to select a representative sample of the Pension and Bounty Land records for testing.

This sample consisted of a variety of document types from items with good contrast to fragile originals with poor contrast. A good mix of page sizes, colors and thicknesses were represented. The average high speed scanning rate for the sample was 31 images per minute. A minimum of contrast level adjustment was used in order to test optimum throughput conditions. Even with little scanning finesse being utilized, the image quality was generally acceptable. Occasionally, an original was in very poor condition making it difficult to read. In these cases, additional image enhancement processing capabilities would have been helpful to produce an electronic image of better quality. Section 6.5.6 on page 117 describes the process and results of using sophisticated enhancement hardware to scan this sample.

6.2.2.5 Government Printing Office Sample

NARA maintains a significant volume of printed documents obtained from the Government Printing Office (GPO). Record Group 287 includes technical manuals produced during and after World War II. Converting these manuals to an alternative media would allow a sizable NARA storage area to be made available for other records holdings. These records were selected to test ODISS's ability to process a collection of 20th century documents, which were relatively consistent in physical attributes such as size and quality. The manual selected for ODISS testing was the *U.S. Department of the Army Technical Manual for Wisconsin Air Cooled Heavy Duty Engines, Instruction Book and Parts List*, Wisconsin Motor Corporation, 1952 (Box 219). This publication has small type sizes, varied fonts, line drawings, halftone images, shaded drawings mostly in black and white on 8.5" X 11" paper.

During high speed scanning of this sample, ODISS processed 36 images per minute. A batch of 21 two-sided documents was scanned two times to verify the throughput rates. These documents were microfilmed, therefore they were sufficiently prepped and provided no handling problems. The documents were of average condition, and the standard threshold scanner settings were used. The 200 dots per inch high speed scanner resolution yielded acceptable image quality when viewed and printed on the ODISS system peripherals.

6.2.3 Scanner Transport Considerations

The only automated document transport system in ODISS was integrated into the Terminal Data Corporation (TDC) high speed scanner. The TDC scanner was modified by the Photomatrix Corporation for Unisys, mainly in the areas of electronic controls and equipment interfaces.

6.2.3.1 Dealing with Different Document Characteristics

The design of the high speed paper transport mechanism was flexible enough to allow for a variety of document types and conditions. Specifically, the transport did not use any type of

grabber or roller to manipulate the paper through the scanner. Instead, an internally generated vacuum holds the paper flat against the transport belts through tiny holes placed between the belts. This technique worked well for virtually all documents used in the test. Light folds and creases were smoothed out and flimsy, thin stock was adequately held down as movement through the scanner took place. Documents such as jackets, folders or envelopes, that are several paper layers thick, benefitted from this vacuum hold-down technique, since the height of the transport zone was not restricted by rollers. Another important factor in the design of the transport was the paper path. The paper moves about two feet in a flat, linear movement until the vacuum is released and then is gently dropped straight down into the hopper. Experience showed that this design was very easy on fragile documents.

6.2.3.2 Use of Polyester Sleeves

Experiments conducted on the use of stock, clear, polyester sleeves for document protection yielded the only poor results to the transport mechanism. The experiments were designed to test folders that were sealed on the various edges. That is, folders were sealed on one edge, two adjacent edges, two opposite edges and three edges. Tests on one sealed edge and two adjacent edges showed that the vacuum would hold down the sealed polyester sides, but was unable to hold the open sides. Polyester sleeves that were sealed on opposite sides or three sides seemed to go through the transport without incident.

During document preparation, placement of an original document inside one of these polyester sleeves was difficult because of static electricity. This static tended to bind the two polyester sheets together and made insertion of the original document time-consuming.

An alternate solution was also tested. Polyester sleeves with two adjacent sides sealed, were used as the basic stock. One sheet of the polyester film was cut on the open side to a size that was approximately .75 of an inch smaller than the other. This enabled the vacuum to hold both the smaller sheet and the overhanging rim of the second sheet. This sealed folder design lent itself to easy document insertion as well.

6.2.3.3 Color Sensitivity

The high speed scanner would produce an adequate image on just about any original document that would fit in the transport. Charge couple device (CCD) scanners are typically "blind" to at least one color. The ODISS high speed scanner could not distinguish certain shades of yellow. This inability had little effect on the CMSR sample although it might for other records. Extensive testing was done by Unisys engineers during system integration to achieve optimum color sensing. An optical filter was installed to produce the best images possible.

6.2.3.4 Sensor Placement

The document transport on the high speed scanner utilized several sensors to determine the proper location of the document throughout the movement of the document. There were sensors along the right side, under the stop lip, that showed when the document was firmly against the right hand stop. This squared-up the document to eliminate any initial skew. Next, there were two sensors perpendicular to the right stop that indicated when the document had reached the forward, start position. When this occurred, the vacuum began and the transport pulled the document through the scanner. Once inside the scanner, there

were other sensors that must "see" the leading and right edge of the paper or a skew condition would be created. The last set of sensors showed when the paper had reached the end of the scan and indicated that the vacuum should be turned off in order to release the document into the hopper.

After some initial adjustment, the sensors generally worked very well for the CMSR sample. One particular annoyance was noticed, however. The upper right corner of the document (upper left corner placed upside down) was very sensitive to edge irregularities. The sensor located there to sense the leading edge of the paper would not turn on the scanner until the paper edge was "seen". If the paper corner was missing, the sensor had nothing to sense. Unfortunately, it was common for the upper left corner of multi-page documents to be damaged or missing because fastening devices once located there had resulted in folding or other damage to the corner during page turning. Therefore, the scanner got a late start when the sensor finally picked up the existing edge of the document. Circumstances such as these had an adverse effect on throughput.

When CMSR jackets were flipped over for scanning (i.e., the bottom is scanned as the first page), the flap of the jacket provided the leading right side. The flap is cut on an angle that causes it not to line up with the leading edge of the jacket (see Figure 6-1). In this case, the sensor had to be fooled by a special operator technique that caused the sensor to think it had the leading corner of the jacket.

6.2.3.5 Other Considerations

Several other important considerations should be mentioned in a discussion of the workings of the scanner transport. A continuous, high speed scanning operation should have two operators to handle the paper transport and several important duties. If the scanned image is to be displayed as it is scanned, someone other than the input operator should be monitoring the screen. This enables the input operator to concentrate on paper placement and pre-placement activities, thereby, increasing scanning throughput. The second operator is also important for unloading the hopper and aiding the input operator when an oversized document needs to be bypassed and placed, in order, in the hopper. As in any conversion operation, operator rotation is also vital to maintain interest and skill level and to decrease boredom.

The tests demonstrated that a scan density of 200 dots per inch was adequate to produce good quality images that were easy to read. An added advantage was that, with such a moderate density, the resulting file sizes were quite reasonable with the average compressed image file of an 8.5" X 11" page yielding somewhere around 40 kilobytes.

6.2.4 Suggested Improvements

The high speed paper scanner, as configured by Photomatrix Corporation, worked very well for the CMSR sample, as well as for the *ad hoc* scanning conducted. There are, however, several improvements that would make the high speed scanning operation run more smoothly and efficiently. These suggestions fall into two categories, systemic and operational.

Systemic changes to the high speed scanner mainly consist of modifications to the existing CPU base. By using Motorola 68030 processors with much faster clock rates, the processing speed of input files could be increased to some extent. Another possible systemic change

Scanner Sensor Placement

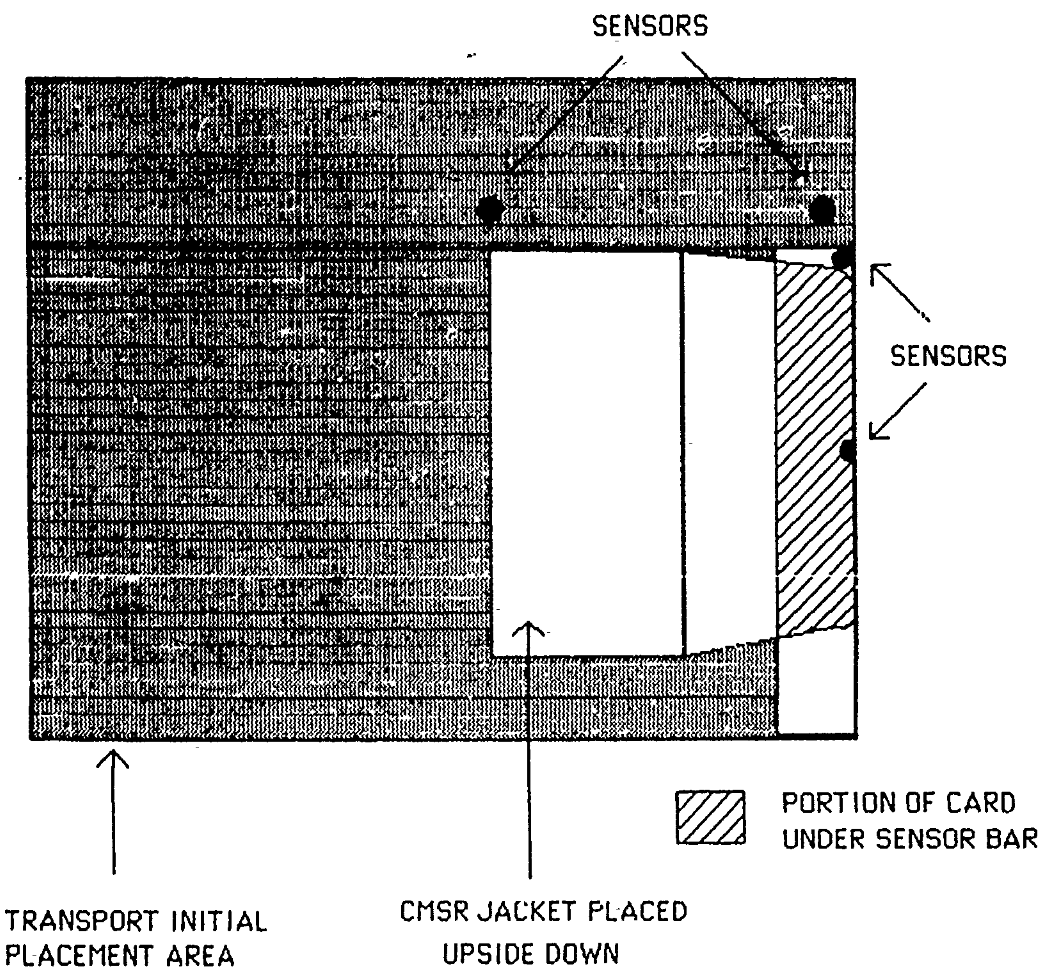


Figure 6-1

would be to utilize completely separate subsystems that handle conversion and retrieval activities.

The ODISS version of TDC's high speed scanner has now been superseded by new models of high speed scanners that have superior capabilities that do not require operator intervention. These new models have sensors that set the size masking automatically. They also utilize dynamic image processing that does not require the operator to manually reset for different document characteristics. If testing with NARA documents confirms these capabilities, current slowdowns caused by manual adjustments and settings would be eliminated resulting in faster scanner throughput.

Operationally, the file sizes could be increased so that the system manager would not constantly have to open and close files at each stage of the conversion subsystem. This could be accomplished even if the files were artificially determined. "Super files," composed of a number of files, could be artificially created in order to expedite transfer through the scanning process. These "super files" could later be broken down into their actual component files.

Another method of high speed scanning input of documents (especially true for irregular documents) would be to utilize a number of table-top scanners and allow the scanner operator to rescan immediately, if necessary, and input index data all in one operation. This would allow more finesse in the scanning process but is not necessary for standard office-type documents.

6.3 Indexing

Evaluation of the indexing functionality focused on ease of use by the operators and throughput rates during the Tennessee CMSR conversion test of full scale production. Information on the ease of learning and using the index workstation was obtained through a questionnaire given to members of the NN input staff after they had several months of work experience in performing the indexing function (see Figure E-1 in Appendix E). Data on the throughput or production rate at the index workstation came from the ODISS management report capability for automatic data collection about all the major ODISS functions. Additional information about the speed of indexing was derived from timings of the index process made by NSZ staff members.

6.3.1 Ease Of Learning the Workstation

The actual work of indexing was typically the quickest and easiest major step in the CMSR input process. Experienced index operators were surveyed for their opinions of the index workstation. They were asked to rate indexing for ease of learning on a scale of 1 to 10 with 1 = easiest and 10 = hardest and they were also asked to select one of five verbal descriptions: (a) very easy (b) somewhat easy (c) average (d) somewhat difficult (e) very difficult. The operators' choices, which are summarized in Table 6-1, indicate that indexing was easy to learn.

Indexing Workstation - Ease of Learning	
Numeric Rating	Number of Operators
1	5
2	1
3	1
Verbal Description	
(a) Very easy	5
(b) Somewhat easy	2

Table 6-1

No operators picked numbers higher than 3, and nearly all picked #1 for the easiest possible. Similarly all the operators picked the verbal descriptions for the two easiest categories.

The operators also were asked how effectively the indexing station performed after they were past any learning difficulties and could operate the station. They rated how well the index station works overall on another 1 to 10 scale with 1 = lowest rating and 10 = highest rating. The ratings were bunched at the favorable end of the scale:

Indexing Workstation - Ease of Use	
Numeric Rating	Number of Operators
10	3
9	2
8/9	1
7	1

Table 6-2

In response to specific questions about different aspects of the index station's functionality, all the operators found the function keys, the numeric code tables, and the default settings (used for the regiment's code value in a run of consecutive files within the same regiment) easy to use.

The operators found the writing and printing on the images usually easy to read but occasionally difficult. When the documents were hard to read, this was most often due to illegible writing on the original documents, but there were some instances where the image quality itself was poor due to either bad initial scanning or a display problem with the terminal's hardware (such as a deteriorating video board).

In summary, the index workstation got high marks and favorable reactions for ease of learning and ease of operation.

6.3.2 Operators' Views

The operators were asked to indicate any problems about matters not covered in specific survey questions and to suggest improvements to the index station.

The most frequently mentioned problem related not to the indexing station itself but to the wait time between files. Slightly more than half the operators felt that the wait time to close one file and open the next was too long and the retrieval of the next file was too slow. This wait between files was a significant cause of the low productivity at indexing and reflected the inability of Unisys's implementation to meet desired throughput speed.

Some of the responses concerned clarifying procedural matters, such as when to use abbreviations in the Remarks field. After the questionnaire was administered, there was some progress toward the standardization of the input for the alphabetical index fields for the names and the remarks. Instructions were developed for punctuation, the use of abbreviations, and the addition of new companies to the code tables. After some months of refining the directions, a formal set of Standard Procedures For Indexing was developed (see Figure E-2 in Appendix E).

Another problem mentioned by a few people was an operator error due to the occasional failure to notice that the regiment had changed so that the default setting for the regiment no longer was accurate; this occasionally caused the wrong regiment code to be assigned to a number of files before the indexer noticed the mistake.

Suggestions for improvements included the system manager's desire to standardize the numeric code tables for companies that belonged to each regiment; this would have avoided some index data entry mistakes and would have saved significant time updating the tables at the system manager's station and then down-loading them to the various workstations.

Two other suggestions related more directly to the indexing station's functionality. One improvement would be the addition of an "insert cursor" to facilitate correcting errors in the alphabetic fields. The second suggestion concerned the rare instance when an indexer noticed a mistake just after hitting the function key to accept the file and retrieve the next file; it recommended a function key to retrieve the previous file similar to the PAGE UP and PAGE DOWN keys used to select the next or the previous page.

These problems and suggested improvements do not detract from the more basic fact that the experienced operators found the CMSR indexing input operation to be an easy and highly rated part of ODISS.

6.3.3 Production and Throughput Rates

The automated management report capabilities of ODISS permitted determining the total production and daily production rates for CMSR files processed during the conversion test for Tennessee records. This test of the feasibility of rapid production using archival records was run between August 8, 1988 and May 26, 1989.

Indexing Totals

The system measured the number of files indexed. The initial workflow plan called for the use of two workstations for indexing and the assignment of other workstations to the function when backlogs developed. Backlogs did occur at various times, and indexing was done concurrently at more than two workstations as needed. Between August 8, 1988 and May 26, 1989, 54,746 files were indexed.

Production Rates

The conversion period for Tennessee CMSR files between August 8, 1988 and May 26, 1989 consisted of 201 available work days. However, daily management reports generated from the automatic data collection by ODISS do not show exactly 201 days of activity at all of the major input functions. For indexing, there were management reports for all 201 days, but three of the reports showed 0.

Daily production can be calculated for either the full 201 available work days or for the number of days worked at each input function as indicated by the management reports. Indexing's daily production rate for the full available work period, i.e., the full 201 available days, is 272 files indexed per day.

If the rate is calculated for only those days showing more than zero, the rate is slightly different. For some days the management reports did show 0; this occurred when a staff member logged onto the system for that function but did no work before logging off. If all the days when no work was performed are subtracted from 201, the result is the number for the days of active operations. Then one can calculate the production rates of each input function as a daily average for the time of active work. For indexing there were 198 active days, and the number of files indexed per active day was 276.

Timings Of The Index Process

A substantial part of the slowness of indexing was due to wait times between files. As the operators noted, there was a varying period of time between completing one file and the arrival of the next file at the index workstation. This wait between files took as much and often more time than the actual work of keying in the index fields.[64]

Timings were taken for small groups of files in early November, 1988, to get some sense of the extent of the wait time situation. The number of timings and the average wait time between files in the four groups are shown in Table 6-3.

[64] The wait times were the result of certain Unisys design decisions during the development of the ODISS system and are not characteristic of digital imaging systems in general.

Indexing Wait Times - November 1988	
Number of Timings	Average Wait Time
8	150.0 seconds
13	43.9 seconds
10	41.3 seconds
25	46.3 seconds

Table 6-3

Three additional timing sessions at the end of December, 1988, and in early January, 1989, still showed long wait times between files. In these three sessions, the average wait time was longer than the average work time, as shown in Table 6-4.

Indexing Wait Times - December 1988		
Number of files indexed	Wait Time	Work Time
45	52 seconds	44 seconds
90	40 seconds	33 seconds
56	56 seconds	32 seconds

Table 6-4

Subsequently the wait time at indexing decreased due to reduced loads on system resources. Timings done in February, 1989 documented wait times in the range of 20 to 30 seconds, and the wait time in these tests still often exceeded the work time spent actually indexing a file. For example, in one timing test of 56 files, the wait time averaged 25.49 seconds and the average work time was 21 98 seconds, while in another test of 58 files, the average wait time was 26.45 seconds compared to an average work time of 24.22 seconds. When file wait times were minimal, index station production throughput rates rose accordingly.

6.3.4 Analysis Of Data

The index workstation was easy to learn and to use. The operators who indexed CMSR files as their daily job gave generally high marks to the station. So, from the standpoint of user friendliness, the indexing workstation was one of the most successful major components of ODISS.

The major complaint of the operators was that the station was too slow in moving between files. The records from timings of indexing groups of files corroborate the operators' criticism. The long wait time between files was also found at other input stations. The slowness of the index workstation is one consequence of the overall sluggishness of the file openings and closings within the input system.

6.4 Quality Control

Quality control was the third major input step in the CMSR conversion process, and it consisted of a 100% review of the Tennessee CMSR files that earlier had been scanned and indexed. Quality control had two purposes. The first was to catch and correct any mistakes made at indexing. The second major purpose was to review the files for image quality and mark poor images for rescanning and image enhancement at the low speed scanner.

Data about the quality control operation were obtained through three methods. A questionnaire (see Figure E-3 in Appendix E) was completed by the experienced operators for information on the ease of use of the station and any problems they encountered during their daily work. Production statistics were developed from the automatic data collection and management report capabilities of ODISS. Timing tests were run to measure how long it took to perform the quality control operation on a file-by-file basis.

6.4.1 Ease of Learning the Workstation

Experienced operators were surveyed for their opinions of the quality control workstation. The survey asked the operators to rate how easy it was to learn the quality control station on a 1 to 10 scale with 1 = easiest and 10 = hardest. The operators also were asked to pick one of five verbal descriptions for ease of learning: (a) very easy (b) somewhat easy (c) average (d) somewhat difficult (e) very difficult. The operators' responses indicated that quality control, although somewhat harder than indexing, was still easy to learn for most people:

Quality Control Workstation - Ease of Learning

Numeric rating	Number of Operators
1	4
2	0
3	3
Verbal description	
(a) Very easy	5
(b) Somewhat easy	1
(c) Average	1

Table 6-5

6.4.2 Ease of Use of the Workstation

The operators also were asked how effectively the quality control station performed after they were past any learning difficulties and were familiar with the operation of the station. They rated how well the quality control station works overall on a 1 to 10 scale with 1 = lowest rating and 10 = highest rating. The ratings clustered at the favorable end of the scale.

Quality Control Workstation - Ease of Use	
Numeric rating	Number of Operators
10	3
9	2
8	1
7/8	1

Table 6-6

The operators also were asked if it was easy to read the index record on the screen, easy to correct indexing mistakes, and easy to use the function keys. Their answers were solicited in the form of choices between "Yes" and "No" with requests for explanatory comments. "Yes" replies mean the person found the activity easy, while "No" responses mean the person had some difficulty. Most operators found each of these three major elements of the quality control station easy:

Quality Control Workstation - Functional Evaluation		
Activity	Yes	No
Reading index record[65]	5	1
Correcting index errors	7	0
Using function keys	6	1[66]

Table 6-7

6.4.3 Operators' Views

The operators were asked to indicate any problems about matters not covered in the specific questions of the survey and to suggest improvements to the quality control station. Some mentioned again, as they had in the survey on the index station, that the system response and wait times were too slow. Some also felt that the table space for working with the paper files was too cramped.

Several noticed that sometimes there were differences in the display clarity of terminal screens. Consequently, images occasionally were marked for rescan at a terminal with poor display quality but appeared perfectly clear and legible when they were displayed at the

[65] The seventh operator answered both Yes and No to indicate that she sometimes had problems reading the images of the jacket.

[66] The one operator who had trouble with function keys sometimes accidentally pressed F10 for "Exit" when he had intended to press F9 for "Rescan."

rescan station. This was not a problem with the functionality of the quality control operation but rather a recurring terminal maintenance and repair concern.

There were some suggestions for changes in the quality control station's functionality. The addition of an insert cursor as a tool for correcting indexing errors was proposed. The ability to remove the rescan mark from an image before completion of the file was suggested by one operator who experienced second thoughts about his judgments on image quality. A couple of operators suggested a capability to change permanently the orientation of images as a convenience for staff and public users on the retrieval side of the system.

Although they pointed out problems and suggested improvements, the experienced operators found the functionality of the quality control operation nearly as easy to learn and to use as indexing.

6.4.4 Production and Throughput Rates

The actions taken by quality control operators at the workstation were recorded by the automatic data collection capability that was part of the management report function of ODISS. The operators compared all the documents in all the files with their images on the workstations' display screens. When the images were hard to read, the operators put the paper document in a colored folder and used an electronic tag on the digital image to mark it as rejected and needing rescanning at the low speed scanner. If the operator found a paper document for which no image existed, the document was put into a colored folder and an electronic "not scanned" tag was placed in the digital file to indicate the need and location for inserting an image at the low speed scanner.

The automated management report capabilities of ODISS counted the various electronic tags for these different actions. So, there were statistics for the number of files approved as having no image quality problems and the files rejected for image quality problems as well as the total images reviewed and the images rejected for poor quality. The system also counted the number of pages in the paper file that the operators marked as not scanned at the high speed scanner.

The figures for entire period of Tennessee CMSR files conversion from August 8, 1988 through May 26, 1989 appear in Table 6-8.

Total Production At Quality Control	
Files Approved	50,152
Files Rejected	8,350
Images Reviewed	256,948
Images Rejected	15,660
Not Scanned	1,336

Table 6-8

The conversion period for Tennessee CMSR files from August 8, 1988 through May 26, 1989 consisted of 201 available work days. However, daily management reports generated from

the automatic data collection by ODISS do not show exactly 201 days of activity at all of the major input functions. For quality control only 199 days of activity were recorded by the automatic data collection facility of ODISS.

The daily averages at quality control for the full 201 available work days are shown in Table 6-9.

Quality Control Production Rates - All Work Days	
Files approved	250
Files rejected	42
Images reviewed	1,278
Images rejected	78
Not scanned	7

Table 6-9

The daily averages for the 199 active days, as recorded by ODISS's automatic data collection mechanism, were:

Quality Control Production Rates - Active Work Days	
Files approved	252
Files rejected	42
Images reviewed	1,291
Images rejected	79
Not scanned	7

Table 6-10

When timings were recorded for files being processed at quality control, they revealed significant amounts of time between the completion of one file and the availability of the next for work.

Some examples illustrate the problem. Two groups of twenty timings were done at two separate quality control stations in late August, 1988, about a month after installation, and the average wait times at the two station were 52 seconds and 49 seconds. Improvements in system performance at quality control are reflected in the shorter, but still overly long wait times found during three tests in late December, 1988, and early January, 1989. The number of files processed and the average wait and work times per files are shown in Table 6-11.

Quality Control Timings - December 1988 and January 1989		
Number of Files Reviewed	Wait Time	Work Time
64	26 seconds	88 seconds
64	28 seconds	47 seconds
58	20 seconds	50 seconds

Table 6-11

By February, 1989, the system performance had improved somewhat to reduce the average wait times found in the next group of timing tests. Work time also was lower in several of these tests as most of the files in this round of tests were much smaller than the earlier ones. Four of these tests illustrate the trend:

Quality Control Timings - February 1989		
Number of Files Reviewed	Wait Time	Work Time
52	19 seconds	22 seconds
51	22 seconds	33 seconds
56	13 seconds	14 seconds
58	15 seconds	18 seconds

Table 6-12

6.4.5 Image Quality Rejection Rate

One goal of the ODISS test was to learn how many images would be judged unacceptable at quality control and require rescanning for better image quality at the low speed scanner. The operators were instructed to mark images for rescan if the printing and writing on the images were not legible.

The overall statistics at quality control show 15,660 images were marked for rescan out of 256,948 images that were reviewed. This is a rejection rate of 6%. So, approximately 94% of the CMSR pages sent through the high speed scanner did not require any further work to make their images more legible.

However, the NN system manager who monitored CMSR production closely on a daily basis felt that there were too many operator errors at quality control. These errors included failing to mark some poor images for rescan and missing extra pages that should have been deleted. This system manager summarized quality control as "Easy to operate, the quality control element has no major technical problems. Poor quality review by operators is the principal deficiency at this station."

The system manager's perception that some unspecified quantity of poor images were not marked for rescan raises the issue of how to standardize the implementation of a quality standard that depends on human review with its potential for subjective differences in understanding even the clearly worded guideline. The instructions were clear and simple, but if the system manager was correct, their application by the operators may not have been as consistent as was desirable.

During the course of the CMSR conversion at the indexing function, more precise standards were developed in response to questions about keying the index fields and these standards ultimately were codified in written form. No similar refinement of a written standard for image quality seems feasible. The standard was simple and clearly worded. Whatever difficulties in adhering to the guideline resulted from variations of human judgment in evaluating images.

Possibly posting copies of good and bad or acceptable and unacceptable images near the quality control stations would help all the operators to follow the instructions in the same way. Technical refinements in such areas as more effective image capture algorithms that adjust better for different document characteristics or the use of techniques similar to the objective tests used in microfilming, such as densitometer readings, might be explored. Further research into both the human and technical dimensions of the definition and implementation of image quality consistency standards in digital imaging systems could be a worthwhile part of a research and development program to follow-up the findings from ODISS.

6.4.6 Analysis of Data

The quality control workstation was almost as easy to learn and use as the indexing workstation, and was a successful component of ODISS in terms of user friendliness.

The major problem noted by the operators and documented in the timing tests was the slowness of the system. In particular, although the problem was not as severe as at the indexing station, there still was a significant amount of wait time between files at quality control, and this wait time substantially impaired the speed and productivity of the quality control workstation.

6.5 Low Speed Scanning and Image Enhancement

The original ODISS system design concept included testing digital scanning's ability to produce quality, readable images from aged, fragile NARA holdings. A low speed scan station for original entry and rescans evolved into a multi-scanner configuration consisting of a Ricoh and a Xerox platen scanner, each with software and hardware enhancement capabilities. Operator selectable menus for scan and image enhancement were provided, with menu-controllable processing of CMSR and non-CMSR files. The low speed scanners also collected production statistics through the system manager.

6.5.1 Gray Scale Image Enhancement Workstation

Unisys originally proposed a low speed system which met all mandatory requirements. Following contract award, in exchange for a concession on an extended delivery date, Unisys provided a refined image enhancement subsystem at no additional cost to the government. A Xerox Inca 8-bit gray scale scanner with 200-400 dpi capabilities was integrated into

ODISS's low speed station configuration. This addition increased the ODISS image processing capabilities, and was useful for two purposes: rescan of extremely difficult documents which required complex enhancement; and as a teaching tool for ODISS operators, staff, and system demonstrations. It was also valuable for determining which enhancement algorithms would be most useful for different types of archival documents.

Station operation involved scanning and viewing selected imagery on a gray scale video display. Enhancements could be performed on the entire document image, or on a smaller "region of interest". The latter is a small area selected from the overall image, which was then expanded to fill the entire display screen. The operator then selected a software enhancement, and viewed the results. Figure 6-2 illustrates the migration from a stained original document, selection of region of interest box, and finally an enhanced digital image.

Although the enhancement station was delivered with eight algorithms for image processing, in practice, not all proved ideal for document processing. This should not be interpreted negatively, since the purpose of this station was to identify which algorithms were most useful for NARA's documents. The algorithms determined as most useful for CMSR documents were: linear contrast stretch, thresholding, and halftone. Linear contrast stretch was most effective for low-contrast documents in which not much visual contrast was apparent between the text and the paper stock. Both dynamic and constant threshold capabilities were specifiable by the operator. These were the most widely used and demonstrated algorithms since they provided the most dramatic "clean-up" of faded, stained documents. Depending on which percentage was selected, the midpoint controlled which pixels would be white and which would be displayed as black.

Optimum selection resulted in images in which the inks/text were retained on a clean white background, with any stains or imperfections removed during the threshold process. In practice, extraneous pixels were randomly scattered throughout the image, requiring removal with "salt-and-pepper" filtering. A histogram display helped determine the optimum threshold points. Histograms were a graphic display of image characteristics, both before and after enhancements. The halftone algorithm was most useful for scanning continuous-tone photographs.

In actual practice, this gray scale workstation was less than ideal for ODISS production operations. Contributing to this was that gray scale imaging greatly increased storage requirements and computer processing times. For example, a single image from an 8 bit-per-pixel/200 dpi scan required 3,200,000 bytes of uncompressed data storage. When the enhancement algorithms were applied to an entire image, or when transferring completed images to the low speed station, excessive processing time was required using the 286-based workstation. It took up to 40 minutes to process fully an 8 bit/400 dpi image.

The elapsed times involved in scanning, viewing, enhancing, and transferring image data using software enhancements were sufficiently long as to preclude the use of this workstation for routine ODISS production. This system was most useful for testing algorithms and demonstrating the enhancement technology to interested observers.

If software-driven image processing were chosen in the future, then workstations with much higher speeds would be required. Considering the success with the Image Processing Technologies (IPT) Scan Optimizer approach (see Section 6.5.3), a hardware-based enhancement process would be preferred because of the reduction in processing times.

Digital Image Enhancement

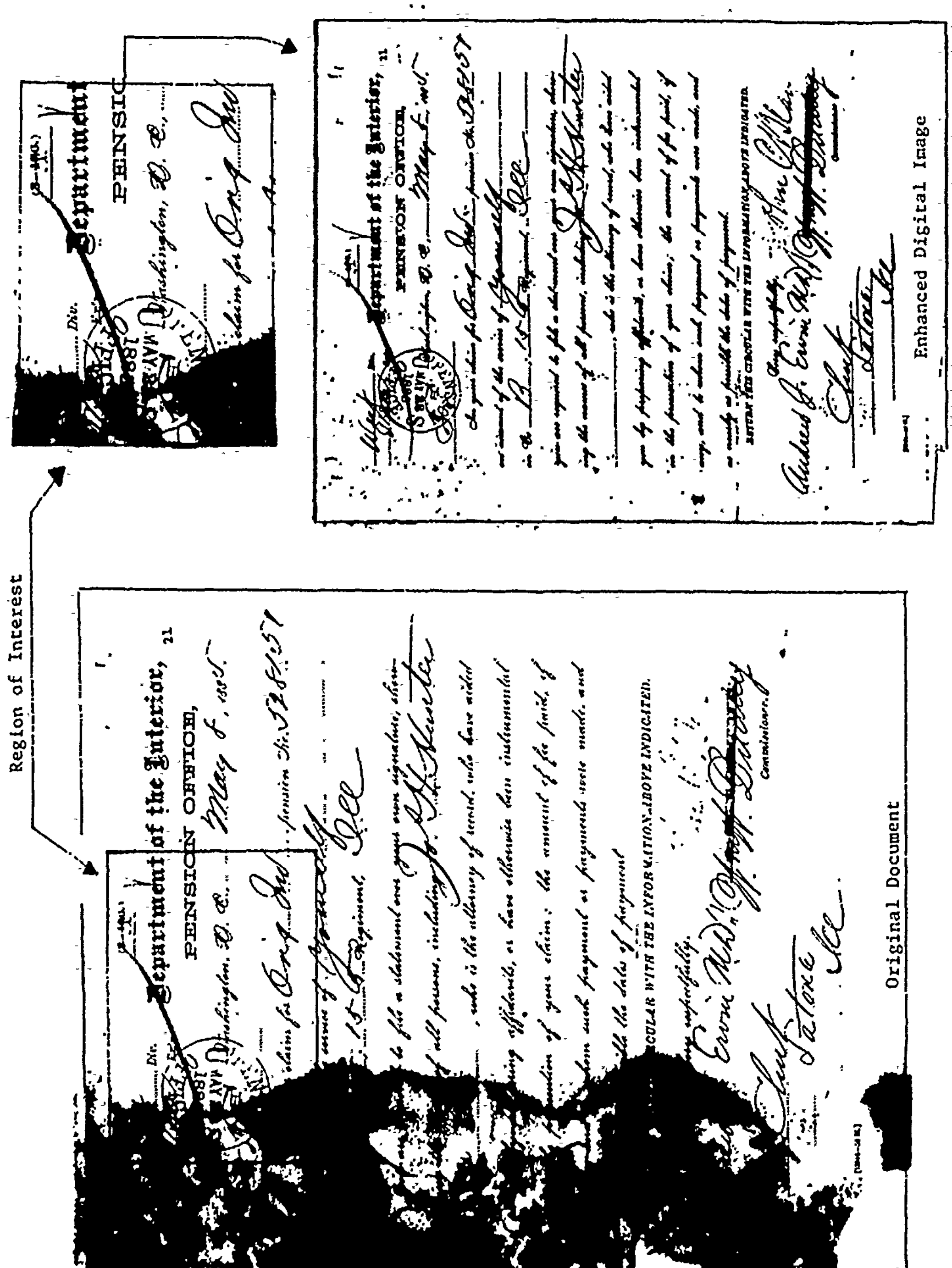


Figure 6-2

Selection of a software-based enhancement process for future production systems would be questionable.

6.5.2 Binary Scanner

A Ricoh IS-400 scanner was originally provided as the primary low speed scanner. Basic capabilities included: variable document input sizes; selectable scan densities of 200, 300, and 400 dpi; binary or halftone scanning; thresholding; and texture removal. These were used by the ODISS rescan operators to control the scanner during original and rescan operations. These capabilities were adequate for many rescanning tasks, although image quality enhancement capabilities were limited.

Station operations were as follows: the operator retrieved the next file marked for rescan, and the operator was then free to replace or insert pages as needed. Rescan or problem pages were scanned and displayed, and quality verified. Image enhancements were made as required, and pages scanned until the desired image quality was obtained.

The preceding actions were repeated until all pages requiring rescan were completed. Also processed at this station were documents with unusual physical characteristics such as fragile, deteriorating paper; edge bindings with several sheets attached; oversized documents; and very faint, low contrast documents in which the original inks had seriously faded. When the operator logged off, a statistical report was sent to the system manager subsystem for future use.

In addition to electronic enhancements, Ricoh scanner image quality was also affected by document physical characteristics and scanner operations. Depending on document construction and appearance, scanned image quality was alterable using various techniques. For example, the low speed scanner's lift-up platen lid featured a white reflective under-surface similar to office photocopiers. This reflective surface often provided superior image quality when scanning high quality documents due to increased contrast between the paper background and textual data. However, when scanning thin, two-sided tissue and other fragile documents, it was often desirable to modify the scan process. Thin, faded documents were susceptible to ink bleed-through during image capture which could, if significant, obscure textual information. This problem was minimized by placing a dark card behind thin documents during scanning, which helped conceal the reverse side data. This process was not an exact science and required operator judgment when selecting the correct background and scanner thresholds.

The low speed scanner was used effectively during the conversion efforts, but ODISS operations staff considered it to be one of the more complex workstations to learn and operate. This is due to the multiple menus needed to navigate the system and the operator judgments and decisions needed concerning scan quality of difficult documents. Optimum low speed station results were obtained when the equipment was used by knowledgeable operators, who were able to select the most effective equipment settings with minimal trial and error.

6.5.3 IPT Scan Optimizer

A new development in digital imaging enhancement products became available after ODISS was operational. Image Processing Technologies from McLean, Virginia developed the IPT Scan Optimizer for installation in document scanning equipment. This product, which was

loaned to NARA on a beta-test basis, was installed in the ODISS Ricoh low speed scanner to provide the manufacturer with field trial experience. The IPT consisted of a replacement scanner circuit board, and an operator control box with touch-sensitive pushbuttons and LED display.

The Ricoh scanner, as delivered, had contrast and threshold imaging capabilities. If a scanned point was brighter than the threshold value, than it was set as a white, or a 1; if not, it was set to black, or 0. This technique was adequate for clean documents; it was not the most effective for seriously degraded documents. The Scan Optimizer, with patented algorithms to adjust a multitude of image processing parameters, improved stained, faded, and low contrast documents. The IPT Optimizer operator could manually set each variable, or use a series of pre-set variable combinations. Operators chose a mid-range pre-set and then adjusted from there as the image warranted.

The goal was to improve document legibility. Ideally, this would also occupy the least amount of digital space by eliminating unnecessary information such as streaks, lines, or noise. The IPT Optimizer used the Ricoh scanner parameters for document size, dots per inch (DPI), and texture removal. In practice, the IPT provided remarkable improvements to image quality and greatly reduced computer processing times. Figure 6-3 shows an example of a low-contrast original document, and Figure 6-4 shows the enhanced image of the document after it was processed by the IPT Optimizer.

Image processing requires a trained operations staff to maximize the quality level of the resulting imagery. Operators' visual judgments are required in conjunction with experience gained in operation of the workstation in order to achieve optimum image quality in a production environment. There are presently no fully automatic image processors that produce optimum results without human intervention.

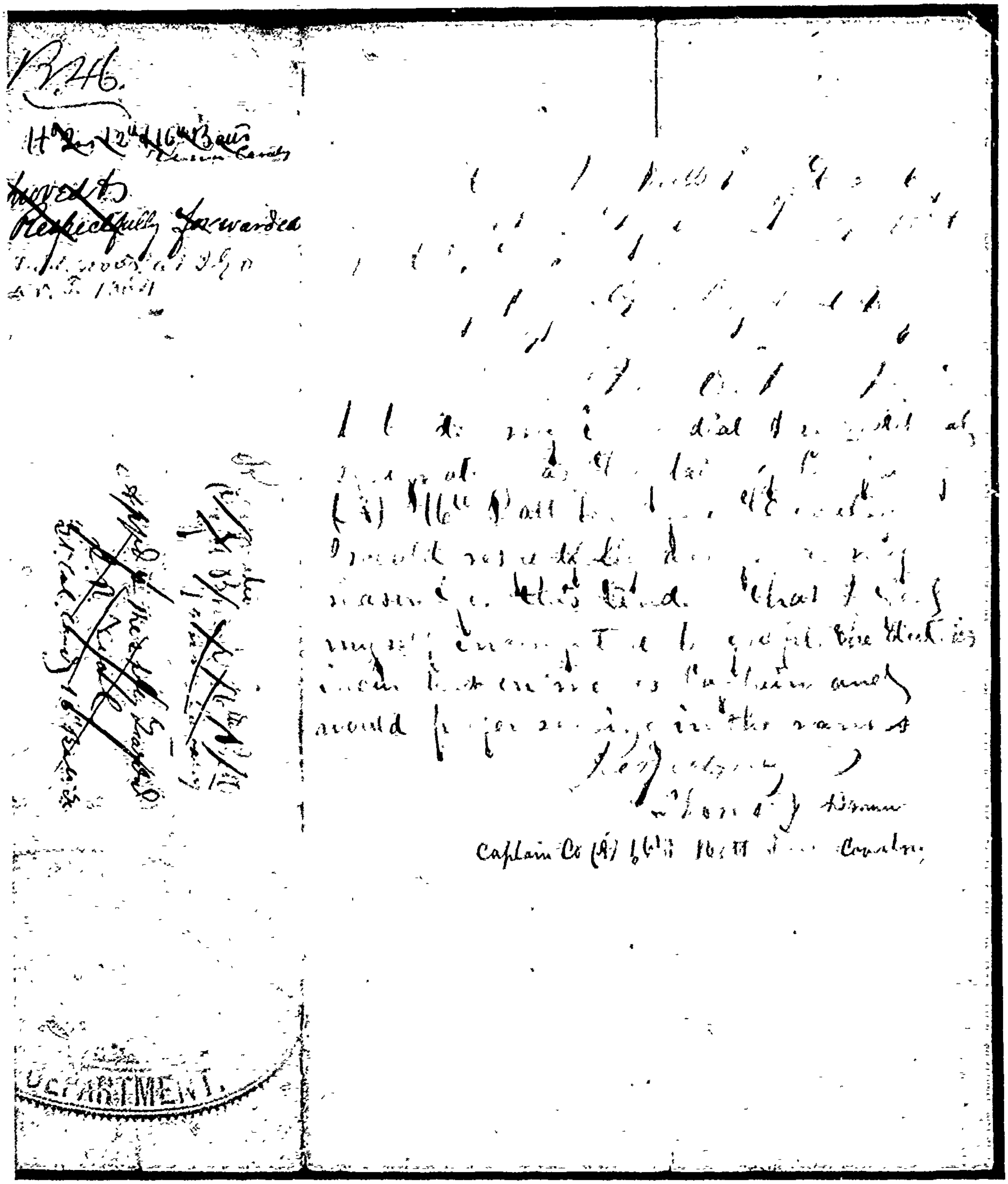
6.5.4 Production Statistics

The low speed station was used to capture original documents that could not be safely sent through the high speed scanner and also to rescan documents whose images captured by the high speed scanner were determined to be of less than adequate quality. Production statistics for the low speed station (not including the gray scale enhancement) during the ODISS Tennessee Cavalry conversion included:

Low Speed Scanner Production	
First-entry files	935
First-entry images	4,266
Rescanned files	6,326
Rescanned images	12,765

Table 6-13

Low Contrast Document



B. 246.

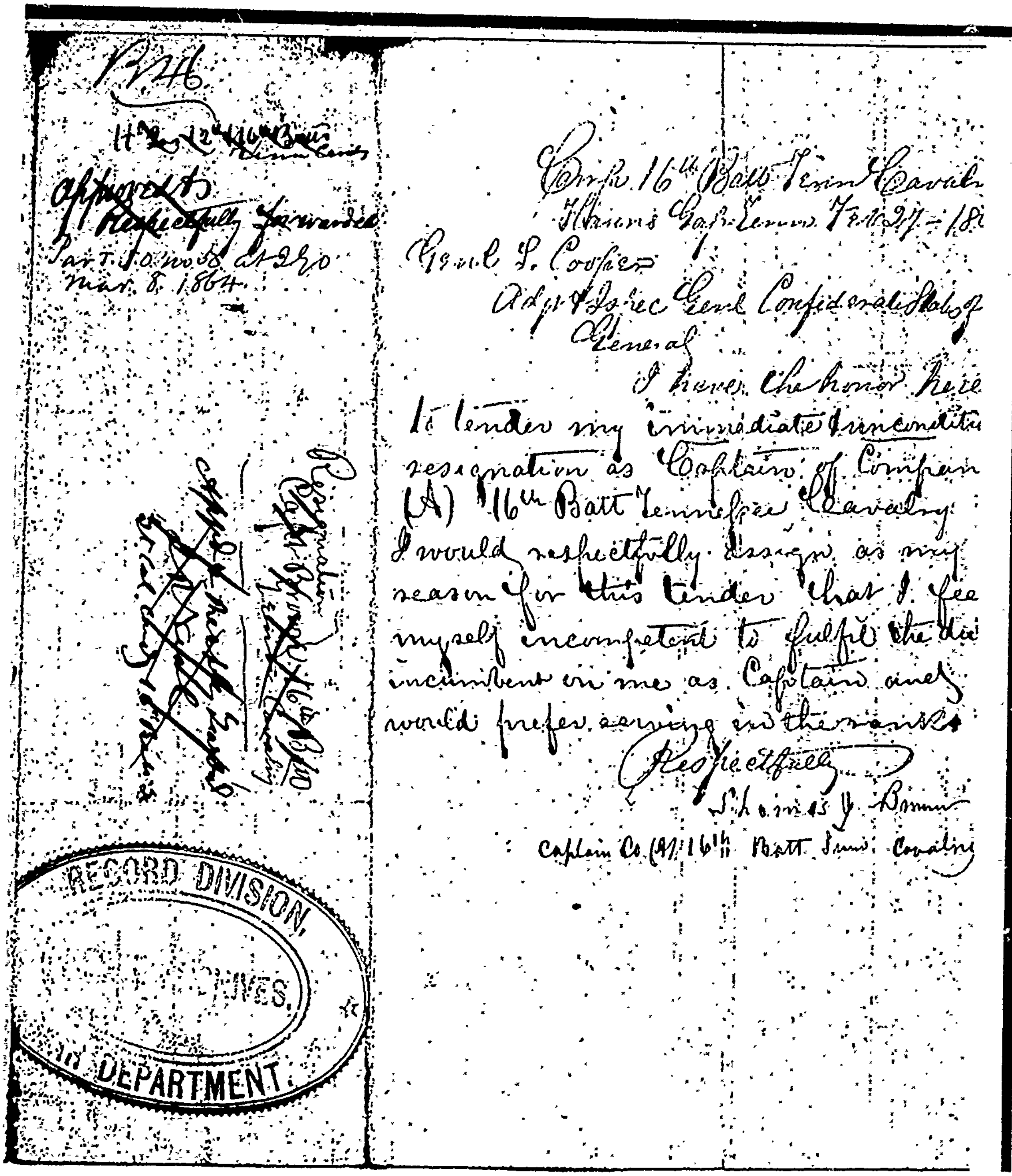
Respectfully forwarded

DEPARTMENT

Captain Co (A) ... Cavalry

Figure 6-3

IPT Enhanced Image



Approved
Respectfully forwarded
Par. T. O. no 58 at 220
Mar. 8. 1864

RECORD DIVISION,
DEPARTMENT

Comp. 16th Batt Tenn Cavalr
Hirams Gap Tenn Feb 27 - 18
Genl S. Cooper
Adjt & Insp Genl Confederate States of
General
I have the honor here
to tender my immediate & unconditi
resignation as Captain of Compan
(A) 16th Batt Tennessee Cavalry
I would respectfully assign as my
reason for this tender that I fee
myself incompetent to fulfil the du
incumbent on me as Captain and
would prefer serving in the ranks.
Respectfully
[illegible]
Captain Co (A) 16th Batt Tenn Cavalry

Figure 6-4

These figures cover August 1988 through May 1989, with an advisory that the statistics are somewhat lower than actual throughput. This discrepancy is due to a data collection problem during several weeks in November 1988.[67] The management reports show that approximately 6% of the file images required rescanning during the Tennessee Cavalry conversion effort. The rescan rate is highly dependent on the skill level of the quality control inspection staff, in that they are responsible for judging screen image quality compared to the original documents. This station required the operators to develop a trained eye for determining whether an image should be accepted or rejected. The most efficient workflow occurred when the quality control and rescan operators were cross-trained in both station's operations. A single rescan station was capable of maintaining production throughput when operated by experienced ODISS staff.

6.5.5 Testing of the Workstation

ODISS low speed scanner capabilities and operational performance tests evaluated system performance and image quality capabilities. Test documents were selected based on a NARA preservation holdings survey, and the test documents were selected based on a cross section of document types.

<u>Test #1 Description:</u>

This test evaluated three areas: compressed byte storage; IPT image quality; and laser printer qualities. Documents from the 1920's were scanned at 200, 300 and 400 pixels per inch. The IPT hardware's ability to improve image quality was examined. Comparison of compressed byte storage was also conducted to determine the impact of higher dpi on storage requirements.

Test Document Samples: Record Group 40, General Records of the Department of Commerce, Employee Reports of Efficiency Ratings were used. Twelve documents, sized 8 x 10.5 inches, from the 1920's were pre-printed forms with typewritten personnel evaluation data.

L/S Scanner Resolution Test: Selected documents verified low speed scanner image enhancement and dpi resolution. Test documents were scanned at 200, 300, and 400 dpi using the IPT enhancement element. The compressed byte storage was verified, and laser prints were produced and evaluated.

L/S Scanner Byte Storage Test: Compressed byte storage was affected by several variables, including dpi parameters, document characteristics, and threshold selection. These test criteria measured the impact of dots per inch resolution on byte storage.

Test Analysis: Scanning at 200 dpi provided adequate legibility for the documents tested. Both 300 and 400 dpi images yielded slightly sharper images, but required more byte storage. Higher resolution images at 400 dpi required more than twice the byte storage of 200 dpi images, and did not always improve image quality. This was due to document imperfections being emphasized at the higher resolution levels. Another finding was that the Ricoh scanner with the IPT processor captured 200 dpi images requiring less storage space than the equivalent high speed scanner 200 dpi images. This was largely due to the cleaner images (less background noise) produced with IPT image processing. Each additional pixel "speck"

---

[67] For details, refer to Section C.3.4 on page 287.

requires storage, even it does not contribute to the information gathering and retrieval process. Laser prints were produced using the ODISS printing equipment for image evaluations. The LP 5400 printers operate at a 400 dpi resolution, resulting in greater print detail than was observable on the 150 dpi display screens.

Test #2 Description:

A series of image quality tests using the low speed scanner determined ODISS's capability with non-CMSR documents. The selected documents tested the scanner's 200-400 dpi resolution capabilities. Image usefulness and compressed byte storage were two important test factors. The tests were conducted under identical procedures with the only differences being the document characteristics. Document selection was based on NARA's document preservation survey, and for this discussion were divided into groups A, B, and C.

Test Document Group A: Twelve documents from RG40, Office of Secretary of Commerce, General Correspondence (Box 172) from the early 1920's. These documents had different colored sheets, tissue paper, dark and light inks, colored ink stamps, and blurred carbon type.

Test Document Group B: Nine documents from RG115, Bureau of Reclamation, General Files 1902-1919 (box 99, folder #127), Unitah Indian Reservation. These documents had faint pencil and light blue annotations, small handwritten notes, ink stamps, black and white photographs, blurred handwritten markings.

Test Document Group C: Thirteen documents from Secretary of Commerce General Office Files, dated early 1900's. These mostly typewritten documents were turquoise carbon ink image on flimsy/transparent paper, purple carbon on brownish stock, light gray carbon, and blue carbon on buff colored paper.

Test Procedures: Test documents were scanned at "optimum" settings using the IPT image processor. Some trial and error was involved to obtain the best initial IPT settings. Test operators experimented with light and dark backgrounds for the thin tissue documents. Single-sided tissue documents required a white backing sheet for best results; a dark backing sheet worked best with two-sided documents to suppress print bleed-through. The documents were scanned at 200, 300, and 400 dots per inch, with IPT settings recorded for future reference. The files were laser printed, and the byte storage for each scanned document was recorded, providing information about the document image compression.

Test Analysis: The original document's physical makeup affected scanned image quality, mandating adjustments to the scanner settings. In general, the 400 dots per inch scan density provided sharper images, especially for areas of fine line detail and character edges. A drawback to 400 dpi use is the unavoidable increase in byte storage. In all cases, document images at 200 dpi scan densities were legible, making the routine selection of 400 dots per inch somewhat questionable. The higher resolutions were usually reserved for special, problem documents which did not produce acceptable results at 200 dpi.

Documents with continuous tone photographs reproduced poorly due to the scanner's binarization process. This indicated that gray scale capability is required if photo content is important. Several documents had a red stain apparently caused by spilled ink. The Ricoh scanner captured the stains's outlined edge, but did not totally obliterate the stain-covered information, resulting in an image that was still useful. Document stains can be removed by the high speed scanner through selective use of optical filters over the capture lens.

6.5.6 Pension and Bounty Land Warrant Sample

The same sample of Pension and Bounty Land records that were used in the high speed scanning tests (see 6.2.2.4 on page 94) were also scanned using the low speed platen scanner. The IPT image enhancement processor was generally left on constant settings so as to achieve as high a throughput as possible. Since the high speed scanner used a scan density of 200 dpi, the IPT-enhanced low speed scanner was likewise set to 200 dpi. Both scanners produced acceptable image quality for the majority of documents from the Pension and Bounty Land Warrant sample, even with the constant settings. With new dynamic enhancement controls (as simulated with the IPT), the image quality would be even better.

As previously indicated, a scan density of 200 dpi produced good, legible images. The use of image enhancement, as in the IPT Scan Optimizer, increases the image definition while producing a much "cleaner" image that compresses[68] at a much higher ratio. This results in an image file that is only about half as large as large as others without this enhancement processing. As seen in Table 6-14, the file sizes with the IPT are generally much smaller than those of the high speed scanner. The documents listed in the table are examples which illustrate the broad range of differences in file sizes when comparing similar scan densities with dissimilar image post processing. The averages at the bottom of the table represent the whole sample of 100 cases (instead of just the 40 cases shown). In every case, the quality of the IPT-enhanced image was greatly improved as well. This sample of original documents required no special handling and was easily converted to fully readable digital images.

[68] Compression removes redundant pixel data such as in a white background (refer also to the Glossary in Appendix I).

Pension and Bounty Land Warrant File Sizes
(in KB compressed)

Document	H/S Scanner	L/S Scanner	Difference
1	45	33	12
2	167	43	124
3	31	44	(13)
4	279	50	229
5	58	51	7
6	47	29	18
7	88	44	44
8	73	45	28
9	64	49	15
10	67	72	(5)
11	84	64	20
12	295	65	230
13	97	64	33
14	179	96	83
15	88	75	13
16	39	26	13
17	96	39	57
18	49	39	10
19	183	101	82
20	81	67	14
21	210	81	129
22	83	69	14
23	75	67	8
24	74	61	13
25	82	56	26
26	99	59	40
27	39	37	2
28	41	44	(3)
29	36	34	2
30	45	31	14
31	33	32	1
33	121	55	66
34	43	36	7
35	69	61	8
36	32	24	8
37	322	91	231
38	80	63	17
39	80	69	11
40	138	91	47
AVERAGES	102	55	47

Table 6-14

6.5.7 Government Printing Office Sample

The low speed scanner with the IPT image processing at 200, 300, and 400 dots per inch scan densities were used for five documents selected from the same technical manual used in the tests of the high speed scanner. Although the documents contained a variety of image features, after some testing and image evaluation, the IPT standard settings[69] were used. Comparison of the 200 dpi compressed storage in kilobytes for the high speed scanner to the low speed scanner is as follows:

GPO File Sizes from High Speed and Low Speed Scanners

Document	H/S Scanner	L/S Scanner
#1	62 KB	51 KB
#2	44 KB	35 KB
#3	64 KB	44 KB
#4	91 KB	68 KB
#5	N/A	83 KB

Table 6-15

This test showed the lower compressed byte storage requirements for an image scanned with the IPT processor compared to the high speed scanner. Lower storage requirements equate to potentially more documents stored per disk. The low speed scanner images are of better quality, due in large measure to the IPT image processor. The IPT images are cleaner (with less background noise), with sharper edge definition on the text and graphics.

The low speed scanner has operator selectable resolution levels of 200, 300, and 400 dots per inch. The same five documents were scanned at the three different resolutions, in order to determine the impact of the dpi selection on the compressed file storage requirements. The sizes of the generated files are shown in Table 6-16.

GPO File Sizes at Different Scanning Resolutions

DPI	Image #1	Image #2	Image #3	Image #4	Image #5
200	51 KB	35 KB	44 KB	68 KB	83 KB
300	77 KB	52 KB	67 KB	107 KB	133 KB
400	114 KB	74 KB	141 KB	151 KB	194 KB

Table 6-16

[69] Standard IPT settings are 8, 3, 7, 3, 0, S, 7, N, and 2.

These results demonstrate the increased storage required at the higher dpi scan rates. The higher values, such as 141 KB at 400 dpi for image #3, were for images which contain halftone graphics. This accounted for the increase at 400 dpi, since a hardcopy of this image contained good image shading and detail similar to the original document. Similar storage rates would be needed to capture extended details of gray scale documents. In comparison, a typical CMSR jacket image required only 32 kilobytes of compressed data to produce a legible image.

6.6 Multiformat Microform Scanner

An ODISS microform scanner made possible a comparison of digital scanning from paper and microforms, and helped evaluate a NARA microform holdings conversion to digital imagery. The film scanner featured format flexibility and multi-level operating capabilities, and was tested with CMSR and non-CMSR microforms.

6.6.1 Operability and Ease of Use of the Workstation

The ODISS film scanner was not utilized for the primary conversion effort; rather, it supported *ad hoc* testing and analysis. The film scanner was incorporated into ODISS to facilitate comparisons of digitally scanned images from paper documents and microforms. Since the Tennessee CMSR records had been previously microfilmed, images from original paper records were compared with images captured from microfilm copies. The film scanner was also tested using various microforms holdings within NARA as well as films from other sources.

Film scanner throughput depended upon film quality and format, image reduction ratio, and image location. The microform scanner's user interface was similar to the high speed scanner, since Photomatrix Corporation developed both devices. The scanner's display screen included image viewing and operator instructions. This scanner accepted 16mm and 35mm roll microfilms, aperture cards, and microfiche from 12X to 48X at six preset reduction ratios.

Film scanner operations involved mounting supply and takeup microfilm reels onto machined spindles, and manual film threading through guide rollers and auto-opening glass flats. The optical-grade glass platen held approximately 5.5 linear inches of film, providing a flat film plane and increased image sharpness. The scanner located the desired film frame(s) by operator-entered coordinates. The operator "eyeballed" the desired film frame, and manually entered X-Y grid coordinates. This process was subject to trial and error, and was not precise enough to guarantee an exactly centered frame every time. In practice, it usually required one additional scan to center the displayed image, or a recalibration to the operator selectable settings. One frequently used method involved scanning at a high image reduction ratio and then viewing multiple images. This was followed by selective rescan enlargement of the desired frame(s).

Any film image located within the glass platen was readily scannable, while other images required platen ejection, manual film advance, and reentry of coordinates. This was a time-consuming process, but had the advantage of accepting roll microfilms lacking "blips" or image count marks. The vast majority of NARA films are not currently blip mark encoded, making image detection more complex.

Station productivity was affected by scan time requirements. After the microforms were loaded in the platen, at least 30 seconds were initially needed to identify manually the grid

location, and key-enter the desired scanner control settings. Following this, single or multiple scans were required. Double scans (12X-16X film reductions) required elapsed times of approximately 28 seconds, while single scans (24X-48X) required approximately 12 seconds scan time. Fine tuning to center the displayed images usually required rescanning using the crosshair positioning software. Selecting the correct enlargement also required some investigation, since many NARA films varied in reduction ratio and were not always clearly identified. Often the operator selected the scanner enlargement after testing selected images, and compared the displayed images to known original document sizes.

Once the desired image was displayed, the operator evaluated image quality to determine legibility. Image contrast and dynamic thresholding were available to improve display quality, with higher quality input microfilms needing only contrast adjustments.

Lower quality microforms usually required the dynamic threshold enhancement to improve legibility. A change to dynamic threshold usually required a different contrast setting. Screen images under dynamic thresholding were not as visually appealing, displaying lower contrast, mottled images with increased "noise". However, under closer inspection thresholded images did provide better edge characteristics and improved legibility. Thresholding could not perform miracles, however; and some of the microforms with critically inferior quality resulted in digital images of marginal quality at best.

Scan density also affected image quality in that higher resolutions usually provided improved screen image display. The 400 dpi scan rate optimized text character edges and also offered an increased image zoom capability when the images were displayed at original scan densities. Optimum digital image quality required a proper balance between contrast, thresholding, and scan resolution.

After image processing activities were completed, the operator could then utilize function keys to manipulate the images using zoom, page rotate, and other capabilities. A demonstration-only windowing capability allowed the blanking of selected portions of screen images. possibly useful in NARA's microfilm declassification applications. Further development of this capability would be required before it is fully useful.

Scanned images could be saved on optical disks and printed on the 400 dpi laser printers. Existing indexing capability on the microform scanner is only non-CMSR, with name and remarks fields available for data entry.

6.6.2 Staff Comparisons of Image Quality from Scans of Paper and Film

Throughout the duration of the ODISS design, integration, factory testing, installation, and conversion activities, image quality was a key consideration. Workstation display screen and print legibility were regarded as pivotal factors in user acceptance of digital imaging technology. NSZ staff continually monitored image quality levels and evaluated staff and casual observers' comments concerning legibility of system output. This area was judged to be significantly important to justify special imaging evaluation sessions with NARA archivists.

A total of sixteen NARA staff from various offices participated in the ODISS image quality evaluation surveys. These sessions were designed to obtain greater insight into the usefulness of digital images for archival applications, and staff member feedback was encouraged. The evaluation sessions were held over two days in the ODISS room using

original documents, ODISS display screens, and hardcopy prints. Prior to the evaluation sessions, ODISS staff prepared test and evaluation samples. These consisted of original CMSR documents and selected pages from Government Printing Office technical manuals containing both text and photographs. These specific samples physically represented the larger body of records from which they were taken, and they also existed in microform.

The evaluation sessions involved visual examination of sample prints produced at 200-400 dpi resolution levels, display image viewing, and staff comments and observations. Hardcopy prints from microforms were also provided for review. All staff members agreed that the 150 dpi screen image displays provided useful images with resolution adequate for archival data retrievals. Laser prints from the 200 dpi high speed scanner were also judged to be adequate. Prints produced from the digitally scanned microforms, as well as conventional microfilm reader-printer copies were also considered acceptable.

6.6.2.1 CMSR Tennessee Infantry Document Tests

In order to expand the evaluation process to other than Cavalry files, Tennessee CMSR Infantry holdings were obtained. Two files selected were Steele, Thomas S; and Steinart, John. Original paper records and a 16mm positive microfilm print were used for these comparative tests.

The Ricoh low speed scanner with Image Processing Technologies (IPT) enhancement capability, and the Photomatrix multi-format microform scanner were used. Scanning was accomplished at 200, 300, and 400 dots per inch resolutions. Digital byte storage levels when available were obtained from the system manager subsystem. The film scanner's 16X magnification was used, and images were not reproduced in exact real size as the original film reduction was approximately 18X. Prior to testing, the 16mm test roll was visually compared to an identical roll in the Microfilm Reading Room to verify microform duplication quality. NARA's Canon NP780 microfilm viewer printers produced hardcopy prints from the 16mm positive microform.

<u>Thomas S. Steele File</u>

Four images scanned from the paper file of Thomas Steele included the CMSR jacket, a card, and two original pay vouchers. These documents were similar to the CMSR Cavalry jackets in appearance and size. The low speed scanner with IPT subsystem yielded the following compressed image sizes in kilobytes:

Image Sizes of Thomas S. Steele File from Paper

DPI	Jacket	Card	Voucher #1	Voucher #2
200	33 KB	32 KB	36 KB	50 KB
300	48 KB	51 KB	56 KB	81 KB
400	66 KB	72 KB	138 KB	95 KB

Table 6-17

A recently produced 16mm microfilm duplicate containing the same imagery was scanned on the multiformat microform scanner with the following compressed image sizes in kilobytes:

Image Sizes of Thomas S. Steele File from Microfilm

DPI	Jacket	Card	Voucher #1	Voucher #2
200	55 KB	N/A	N/A	91 KB
300	92 KB	N/A	N/A	128 KB
400	132 KB	N/A	N/A	168 KB

Table 6-18

Byte storage for the card and first voucher [listed as N/A] in Table 6-18 could not be accurately determined. This was due to minimal spacing between these two microfilm frames, resulting in the two images being scanned as one frame.

John Steinart File

A second file was selected, containing three John Steinart documents including a CMSR jacket, a card, and letter on blue colored paper. The Ricoh document scanner generated image file requiring storage amounts as follows:

Image Sizes of John Steinart File from Paper

DPI	Jacket	Card	Letter
200	31 KB	32 KB	49 KB
300	47 KB	49 KB	80 KB
400	85 KB	73 KB	211 KB

Table 6-19

The identical images were scanned from the 16mm microfilm, and microfilm reader printer copies were produced for visual comparisons. The image file sizes from microfilm are shown in Table 6-20.

Image Sizes of John Steinart File from Microfilm			
DPI	Jacket	Card	Letter
200	62 KB	N/A	79 KB
300	102 KB	N/A	123 KB
400	149 KB	N/A	175 KB

Table 6-20

Images from both the document and film scanners were legible at scan densities of 200-400 dpi, although as expected, sharper character edges were attained with 400 dpi. The low speed document images were of slightly higher quality than the same images scanned from the microfilm. This resulted from the document scanner working with the original documents, while the film scanner used a successive generation microfilm copy.

Microfilm scanned images required almost twice the compressed byte storage than those from the original paper records. This was due to the increased compression offered by the Ricoh electronics and the fact that the film scanner captured dust, debris, and microfilm defects. Also, the microfilm images had higher density (darkness) leading edges, probably due to uneven document exposure during the original microfilming. This density area, which is observable on a film reader, required digital storage even though it was non-data. The IPT is a powerful enhancement capability which could probably improve the film scanner's image quality and lower the requirement for compressed image storage. This could not be verified without integrating the IPT processor into the film scanner.

Comparison of digital laser prints and microfilm viewer printer hard copies was conducted. Producing hard copies from microfilm was cumbersome because NARA's Canon NP780 viewer printers had 12X or 24X lenses. These did not match the 18X microfilm, as prints were either smaller or larger than the original 9.5-inch-high jackets. This mismatch is an ongoing problem for any microfilm in the Microfilm Reading Room which is not 12X or 24X.

NARA has a Minolta microfilm reader printer with a 17X lens, but this machine is an outmoded electrostatic wet process offering low contrast prints not considered useful for these tests. Prints made directly from the microfilm were compared to those produced after digital film scanning, and samples were viewed by NARA staff as part of the image quality analysis. NARA staff consensus was that both the digitally scanned and conventional microform images were fully legible and useful for the intended purpose of information retrieval.

6.6.2.2 Government Printing Office Document Tests

NARA is currently converting GPO records holdings to industry-standard 24X, 98-image microfiche with eye-readable titles. Sample documents and silver halide master microfiche were selected from the GPO records holdings. Since the GPO conversion is using current NARA processes and quality control procedures, the GPO microfiche were of consistently high quality. The documents selected were U.S. Dept. of the Army technical manuals for Wisconsin Air Cooled Heavy Duty Engines, the same as used for the high speed scanner tests described on page 94 in section 6.2.2.5.

The silver halide camera master microfiche selected were evaluated for image characteristics. The original documents contained engine parts illustrations in various shades of gray. These were captured on the microfiche in higher contrast, resulting in loss of some tonal image data prior to digital scanning. Due in part to this, the digital images scanned directly from the documents were judged to be of slightly higher quality than images from the microfiche. Images of microfiche 03509 were digitally scanned, stored on magnetic disk, and printed on the ODISS laser printers. These same images were also printed on NARA's Canon NP780 microform viewer printers.

The test results indicated that a film scan at 200 dpi is adequate for most reasonable quality microforms, although the addition of an IPT capability to the film scanner would probably improve the overall image quality and reduce the storage requirements. The film scanner produced sharper images at 300 and 400 dpi, although at the base 200 dpi, the images were legible.

After scanning, hardcopy prints produced on the ODISS laser printers were compared to Canon microfiche reader-printer copies. The main difference between the 200 dpi scanned film image prints and those produced by the microfiche reader-printer is character edge definition. The laser prints from scanned film images have slight jagged character edges, resulting from the scan process. These characteristics were not detrimental to print legibility, and were not obvious to the casual observer. Both sets of test prints were judged by many observers to be fully useful for information retrievals.

6.7 Optical Storage and Archiving

Optical disk storage was a pivotal technology within the ODISS project. Its use of laser energy to alter the reflectivity of a disk surface at infinitesimally precise tolerances was in itself an impressive feat. As applied to ODISS, the optical disk drives used to store and retrieve raster images produced in a routine production environment worked as advertised in that they performed reliably with no significant downtime or loss of information. The jukebox retrieval subsystem consistently delivered images within the twelve-second response time requirement. This section further discusses the archival workstation operations and performance experienced during the ODISS conversion activity.

6.7.1 Archive Process Overview

The archive process was the last step in the ODISS conversion cycle of creating digital CMSR images. After blocks of files were scanned, indexed, quality control reviewed, and rescanned if necessary, they were transferred to permanent optical disks. Writing image data to optical disks is called, in computer jargon, "archiving." In ODISS, this archiving process permanently stored images that had been temporarily maintained on magnetic disk. As the images were recorded on the optical disk, the database was updated through the system manager to change the images' location from magnetic to optical disk. This database update meant that the images could be subsequently accessed only from the optical disks. The images were erased from the magnetic disks which were then made available to store newly scanned images. This in-process workflow was actively monitored by the ODISS system manager. More detail on these hardware processes is provided on page 245 in section B.7.3, Archives Subsystem.

6.7.2 Archives Workstation

Optical disk writing was performed using the system manager's terminal under a special archive function. Archiving was performed by the NN staff member trained as the system manager, augmented by members of the NSZ ODISS project staff. The archive operation was controlled through a series of system manager menu choices. Prior to archiving, printed reports of the batch file contents were created to facilitate management oversight of archive operations for permanent retention.

The System Manager workstation's main menu offered nine operations, and after selecting archive management, a four choice menu (Figure 6-5) appeared. Choice #2 listed blocks ready for archiving (Figure 6-6). After determining which blocks were available, the operator returned to the Archives Management menu and selected choice #4. The Initiate Archive Process menu (Figure 6-7) appeared with prompts to enter the block to archive, and the optical disk side and volume number to write the images. The screen then presented the question, "Initiate Archive (y/n) ?" (Figure 6-8) to which the operator would select "Y" followed by a carriage return. The archive process would then begin.

ODISS archiving procedures utilized the outboard drives and controller. The disk was loaded prior to beginning the write process, with the write-protect safety mechanism deactivated. The write-protect was a manual slide mechanism, which was tested and determined to be effective in preventing unintended disk modification. Once the write process was started, the Archives Management menu allowed progress monitoring (Figure 6-9). The Archive Management menu listed the available optical disk space in the system (Figure 6-10), useful when archiving new images. The archive menus were easy to learn and use. In the words of the NN system manager, "The Archive Management functions have proven very satisfactory."

No operational problems were noted during the initial archiving attempts, and Unisys instituted a pre-archive procedure to check data blocks for potential problem attributes. This software procedure examined the blocks awaiting archive, and checked them for outstanding or incorrect codes within the image files. These incorrect codes resulted from system errors generated during routine workflow processing, and would halt the archive process prematurely. This menu driven pre-archive function occupied an ODISS workstation when operated, and successfully corrected the archive processing problem.

6.7.3 Optical Disk Security Backup System

Security copies of the original optical disks were created by ODISS. This helps to avoid the potential costs involved in reconverting original documents if the original disks are damaged. ODISS originally was to backup image data using magnetic tapes on reels, but the potential tape quantity led NARA management to identify a better approach. This included backing up optical data onto duplicate optical disks. To accomplish this, two optical drives were required, with software to initialize the backup process.

ODISS used Sony's twelve-inch, two-sided constant angular velocity (CAV) optical data disks. The disk capacity was 1.091 gigabytes[70] of user data per side, or 2.18 gigabytes total.

---

[70] One gigabyte is one billion (1,000,000,000) characters.

Archive Subsystem Screen: Menu

```
[archive]                                          29 JAN 1988 - 09:48

              NATIONAL ARCHIVES AND RECORDS ADMINISTRATION

                      ODISS ARCHIVE MANAGEMENT

                 1.  Display Status of Previous Archive
                 2.  Blocks Ready to Archive
                 3.  Optical Disk Free Space
                 4.  Initiate Archive Process

SELECTION:

ESC-select   ^U-up   RET-down   ^X-home   ^P-previous   ^Z-clear
^D-exit   ?-help   /-more
```

Figure 6-5

Archive Subsystem Screen: Ready to Archive

NATIONAL ARCHIVES AND RECORDS ADMINISTRATION

BLOCKS READY TO ARCHIVE

STAGE = ARCHIVE STATUS = CLOSED

---

1 2

Figure 6-6

Archive Subsystem Screen: Initiate Archive Process

NATIONAL ARCHIVES AND RECORDS ADMINISTRATION

INITIATE ARCHIVE PROCESS

STAGE = ARCHIVE   STATUS = CLOSED

1   2

Block: Side: Volume:

Please enter block, side, and volume

Figure 6-7

Archive Subsystem Screen: Initiate Archive? (Y/N)

NATIONAL ARCHIVES AND RECORDS ADMINISTRATION

INITIATE ARCHIVE PROCESS

STAGE = ARCHIVE   STATUS = CLOSED

1   2

Block: 2   Side: A   Volume: 999997

Initiate Archive (y/n) ?
Please enter block, side, and volume

Figure 6-8

Archive Subsystem Screen: Progress Monitoring

```
            NATIONAL ARCHIVES AND RECORDS ADMINISTRATION
                           ARCHIVE LOG

Friday Jan 22 09:39:36 1988
Archive of BLOCK 4 initiated : there are 10 FILES

==========================================
Request-to-Archive Message Written
Archive Response Message Read
CMSR record updated
ODISK record updated
FCNBLOCK record deleted                  fcn=90224
1 of 10 FILES archived
------------------------------
Request-to-Archive Message Written
Archive Response Message Read
CMSR record updated
ODISK record updated
FCNBLOCK record deleted                  fcn=90225
------------------------------
f=forward               Choice:                        q=quit
```

Figure 6-9

Archive Subsystem Screen: Available Disk Space

NATIONAL ARCHIVES AND RECORDS ADMINISTRATION
OPTICAL DISK FREE SPACE

NODE	VNUM	SIDE	% FREE	NODE	VNUM	SIDE	% FREE
A	000000	A	99	A	000000	B	100
A	000001	A	99	A	000001	B	100
A	000002	A	99	A	000002	B	100
A	000003	A	99	A	000003	B	100
A	000004	A	99	A	000004	B	100
A	000005	A	99	A	000005	B	100
A	000006	A	99	A	000006	B	100
A	000007	A	99	A	000007	B	100
A	000008	A	99	A	000008	B	100
A	000009	A	99	A	000009	B	100
A	000010	A	99	A	000010	B	100
A	000011	A	99	A	000011	B	100

Choice:

---

f=forward q=quit b=backward

Figure 6-10

Once filled, a backup optical disk copy was made. Odd numbered disks were considered masters, and the even numbered disks were the security copies. Under this procedure disk one is an original, disk two is the backup copy of disk one, disk three is the second original, disk four is the backup of three, etc. Backup copies were remotely stored in case of a problem in the immediate production area. This ensured that the data survives a disaster in the main image processing area.

Also requiring permanent retention was the key-entered index data, which remained on magnetic disk drives. Security backup copies of index data were created daily on streamer tape cartridges. During the ODISS conversion effort, neither the magnetic streamer tapes nor optical disk back-up copies were needed for any recovery effort.

6.7.4 Optical Disk Longevity

Obviously, the ODISS project had no way to verify the long-term longevity of optical disk media. Longevity is, however, a very important aspect of any electronic storage medium. Questions related to optical disk longevity arise such as, how long will the data safely reside on the medium and still be retrievable? Or, can the operator determine when the data is just beginning to degrade? And, can the data be copied to another optical disk (or another digital storage medium) with no loss of data from generation to succeeding generation?

To address the data disk longevity question, NARA helped to establish a National Institute of Standards and Technology (NIST) laboratory for independent determination. NIST staff will develop a standard testing methodology to verify vendor longevity claims. Currently, some optical disk suppliers are quoting one hundred year life expectancies, and are privately saying that they really expect five hundred years with proper disk storage. Media life is important since optical disks are not "human readable". Optical disk data is compressed and not interpretable without a compatible computer system and optical disk drive. Therefore, system life is linked to the disk media. Optical disk systems offer early warning indicators for impending degradation. By periodically verifying specific disk sectors, and comparing the results to pre-established benchmarks, data integrity can be assured.

One significant benefit of digital storage media is data transportability. If an existing disk begins to deteriorate, data can be copied onto another optical disk, magnetic tape, or any other digital storage medium. No original data is lost, and no document rescanning is needed. With the constant price decreases for optical storage media, coupled with higher performance, some type of recopying schedule should be evaluated. This would allow a system to remain up-to-date with this very dynamic technology.

6.7.5 Analysis of WORM Disk Capacity

One major advantage of optical disk is its very large capacity for data storage. Industry claims for write once, read many times (WORM) optical disks are that they are capable of holding much more data than alternative media. NARA projections for the ODISS documents were 20,000 compressed images per side, or 40,000 document images for the two-sided disks. These projections proved to be fairly accurate.

The Tennessee cavalry conversion required five complete optical disks, with part of a sixth disk containing 2,275 Tennessee CMSR files with 6,182 images. Statistics for the five completed disks are shown in Table 6-21.

Optical Disk Utilization	
Disk #	Images on the disk
1	41,422
3	38,273
5	43,130
7	44,097
9	47,609

Table 6-21

The average disk capacity was 42,906 CMSR images, which exceeded initial projections by about seven percent. The Tennessee CMSR cavalry files, previously stored in 946 archives records boxes, now occupied little more than five optical disks. Capturing the Tennessee Cavalry on only five disks clearly demonstrates that optical disks are capable of storing archival records in remarkably reduced space. Table 6-22 provides a comparison of storage reductions possible with microfilm and optical disk technologies for holdings of a quarter million up to a half billion documents. For example, the fifth row of the table shows that documents comprising 50 million pages would occupy 20,000 cubic feet of shelf space. By comparison, images of the same number of pages could be stored on optical disks measuring 25 cubic feet in volume.[71]

Comparison of Storage Requirements			
Pages	Paper	Microfilm	Optical Disk
250,000	100	3.47	0.13
1,250,000	500	17.36	0.63
2,500,000	1,000	34.72	1.25
12,500,000	5,000	173.61	6.25
50,000,000	20,000	694.44	25.00
250,000,000	100,000	3,472.22	125.00
500,000,000	200,000	6,944.44	250.00

Table 6-22

6.7.6 Operational Experiences

The archive subsystem's electro-mechanical and optical components performed as expected, with minimal downtime following correction of initial development problems. One minor issue noted during testing was print request service procedures versus optical drive

[71] The figures shown in Table 6-22 are based upon the use of 35mm roll microfilm and 12-inch optical disks capable of holding 80,000 document images per disk. The figures only include space requirements for the storage media. They do not include space required for workstations or retrieval devices (e.g., jukeboxes).

allocation. That is, during a routine search the jukebox would fetch a disk and load it into one of the available resident drives. After transmitting the images to the display screens, the jukebox would return that disk to its assigned shelf location. If the user requested a print from this same file being viewed, it was handled as a totally new transaction. The disk had to be fetched again, and reloaded into the drive. Some loss of efficiency resulted from this operational design.

6.8 Staff Retrieval

Research tests were performed to collect data on the use of the staff retrieval workstation by the NARA staff. The tests were designed to collect information on both the ease of using the staff retrieval functions and the speed at which searches could be performed.

6.8.1 Test Design and Procedures

The research test was designed for NARA staff members who already knew how to perform searches of the Confederate CMSR records with the original records. These staff members would learn how to use the staff workstation. Then they would perform searches for Tennessee Confederate CMSR cavalry files that are stored in ODISS. They would record the time it took to perform the searches and the results of the searches on a form; this would allow comparison with the speed in the present manual system of searching for the original paper files. After finishing the search batches, the staff members would be asked to complete a questionnaire about the various phases of learning and using the staff workstation. The questionnaire would allow the staff members to summarize their experiences and judgments on the staff workstation.

Prior to the testing sessions, NSZ prepared six batches of ten inquiries each to simulate the batches of mail-in queries that NARA staff members answer. The batches consisted of a mixture of easy and difficult queries. NSZ also drafted a "CMSR File Search" form for the searchers to record the results of each of the ten searches in each batch as well as the starting and stopping times for processing the batch. Like the standard NNRG-P batch forms, there were also places on the form for a reviewer to record the error rate in the searches and the total time spent processing the batch. For the final questionnaire NSZ developed thirteen questions in a "Staff Reference Data Collection Form."

Figure E-4 in Appendix E is an example of the search batches. The first and last pages of the "CMSR File Search" form are shown in Figure E-5. The three-page final questionnaire appears as Figure E-6.

6.8.2 Test Implementation

Four NNRG-P archives technicians whose regular duties include making searches of Confederate CMSR records for replies to mail-in queries were selected to participate in the test. The test was conducted in two separate single-day sessions, each of which involved two of the NNRG-P participants.

The technicians were introduced to the staff workstation by an ODISS project team member from NSZ. The station's menus and function keys were demonstrated. All major functions of the staff retrieval workstation were explained; these included conducting searches of various levels of difficulty, working from index search "hit" lists, retrieving and viewing file

images, paging through files, rotating and zooming images, and printing both index lists and file images.

The NNRG-P technicians then had some time for hands on practice in using the major functions of the workstation. The NSZ staff member was available to help them with problems in learning the system. They had approximately one hour of practice time before the next step in which the batch search process was described, including filling out the search form. The process was demonstrated by doing all the searches in one batch and filling in the form as the technician completed the ten searches. After this demonstration, the NNRG-P participants were given more practice time for about one hour.

Then the test participants were given batches to process on the staff workstations and, as they performed the searches, they recorded their work on the "CMSR File Search" forms.

In the final phase of the test sessions, the NNRG-P staff members completed a questionnaire that asked them about the ease of learning and using the different facets of the staff workstation.

As administered, the test sessions had some limitations. The major limitation was the small amount of practice time available for the NNRG-P participants before they had to perform the search batches. The hands-on training time totaled only about two hours. Because of the differences among the participants in speed of learning a completely new system there were some differences in the ease with which the four people used the workstation after such a short learning period. These differences are reflected in the fact that some participants did a single search batch while others did two separate batches.

6.8.3 Ease of Learning and Use of the Workstation

All four NNRG-P staff members thought the staff workstation was easy to learn. Three gave a rating of 1 and the other person gave a rating of 3 to the question "How easy or hard was it to learn to operate the staff workstation on a scale of 1 to 10 (1 = easiest and 10 = hardest)?" Another question that asked people to select a verbal description of how easy or hard it was to learn to operate the staff workstation gave five choices: very easy, somewhat easy, average, somewhat difficult, and very difficult. Three of the participants picked "very easy," and the fourth chose "average."

All four test participants perceived the workstation to be easy to operate after their initial learning phase. In reply to a question asking, "After you have learned to operate the staff station, how well overall did the station work on a scale of 1 to 10 with 1 = lowest rating and 10 = highest rating?" each person chose the rating of 10. All four also reported that using the image rotation and zoom features, employing the code tables to build searches, shifting between index hit lists and file images, and printing copies were all easy. One of the technicians said that getting files on ODISS was better than having to get files off the lower shelves in the stacks.

When asked if the function keys were easy to use, all four said yes, and none selected any of the options from a checklist of possible problems that was provided on the questionnaire. In response to a question asking, "Is the writing/printing on the images of the CMSR files easy to read on the screen?" three people picked "always easy" and one picked "usually easy."

The last question on the questionnaire asked, "Compared with the current way of servicing CMSR records how do you rate the ODISS method on a scale of 1 to 10 (1 = lowest rating & 10 = highest rating)?" One person picked 6, explaining that, "I can do the current way fast;" one chose 8; one picked 9; and one chose 10, adding the written comment that, "I think the system is fast and very clean. I think we should keep it."

6.8.4 Production Rate

During the test sessions, the NNRG-P archives technicians completed six search batches. However, one person forgot to record the time for a batch. Consequently, there are five batches for which the staff searchers recorded both the starting and completion times. Since each batch contained ten searches, an average time per search in each batch can be calculated. The total times and average times per search in minutes for the five batches are shown in Table 6-23.

Staff Search Time Test

Total Time	Average Time Per Search
17	1.7
17	1.7
28	2.8
34	3.4
45	4.5

Table 6-23

In the current system for CMSR reference service using the original paper records, the existing performance standard to complete a search is 9.6 minutes, while the time to exceed the standard is a maximum of 7.8 minutes. All the times of even the slowest searcher on the ODISS staff workstation were substantially faster than the times required in the current system.

6.8.5 Search Accuracy

The accuracy of the searches was affected by the limited training and practice time that the participants had to learn the system before they worked on the search batches. Their total "hands-on" practice time was only about two hours.

Moreover, the rules for a successful search in ODISS reflected the capabilities of the automated database to support more detailed and flexible searches than may be conducted in the current system. Some of the NNRG-P technicians said that only the exact name given by the requestor is searched, while a number of the searches in the test batches on ODISS required using the rapid retrieval ability of the database to look under variations of the subject's name. So, while a search in the current system might stop at whatever results were found under an exact first name, last name and no middle name, the ODISS search capability took into account the possibility of various middle names or initials.

The rules for searching on ODISS required following up cross referenced variations on a name in the index's Remarks field to locate second files that are generally larger. The largest single kind of errors in the search batches resulted from failing to use the cross references in order to get to the second file under a variant spelling of the name that actually had most of the records on the person.

Despite these problems, the accuracy rate for the search batches included some good results. Moreover, the accuracy did not correlate precisely with the speed of the searches. The accuracy rate is calculated as a percentage of the total searches; if the person got seven correct out of the ten in the batch, the rate is 70%. The rates for all six batches with the times in minutes are shown in Table 6-24.

Staff Search Accuracy Rates

Total Time	Time Per Search	Accuracy Rate
not recorded	----	50%
17	1.7	50%
45	4.5	70%
34	3.4	70%
28	2.8	80%
17	1.7	90%

Table 6-24

In the one-day test sessions, there was not enough time for all the technicians to learn thoroughly the broader, more flexible rules for performing searches on ODISS. This was reflected in some comments made to the NSZ staff member conducting the test sessions. Most of the NNRG-P participants said words to the effect that, given a week to use the system, they believed that they would have so mastered it that they would have been able to work on it with complete skill and confidence.

6.8.6 Analysis of Test Data

The results of the test sessions with the four archives technicians who perform searches of the Confederate CMSR paper files indicate that performing the same kind of work for digital versions of the CMSR records at staff workstations is feasible and, in fact, could be very beneficial to NARA.

The tests demonstrated that it is feasible to teach the current staff as well as others of the same grade level and similar backgrounds in terms of skills how to use a computer terminal to perform the CMSR search activity. All the participants felt that they were able to learn to operate the staff workstation, that they picked up many of the basics in the short practice time of about two hours, and that with a few days of experience they could become totally proficient.

The accuracy rates in the test's search batches give objective support to the participants' perception that they were mastering this entirely new activity. Four of the six batches had

accuracy rates of 70% or better, and two of the six were in the 80% to 90% range. The worst rates were no lower than a 50% accuracy level.

The most dramatic result of the test was the speed at which the archives technicians were able to complete their searches. The slowest rate was approximately twice as fast as the acceptable rate in the current system, and the fastest rate of a little less than two minutes per search was about four times as quick as the present system.

These results may need some qualification. The test search batches were ten requests rather than the 57 queries in NNRG-P's standard Confederate CMSR batch. There is a legitimate question whether the same fast rate could be sustained with larger batches and as a full time, everyday job. However, the test results still indicate that much greater productivity would be possible in an automated, digital image-based reference system.

The benefits of this greater productivity could accrue to both NARA as an institution and the employees performing the reference function for mail in CMSR requests. Quicker response times might mean better public satisfaction with NARA's service, and greater productivity would improve NARA's use of its own resources of equipment and staff. The staff performing the CMSR reference function might benefit both in pride in using a new technology and in financial rewards for greater productivity.

In summary, the tests of the staff workstation indicate that this aspect of ODISS worked well. The staff workstation received favorable responses from the archives technicians most familiar with Confederate CMSR reference searches. They felt that they could learn to operate the workstation skillfully within a reasonable time. The test results for both speed and accuracy obtained from the technicians with only minimal learning and practice time supports the view that the staff workstation can be an effective system for NARA employees. The results on search speed clearly indicate that an automated index and digital image system would be much more productive than the current manual, paper record system for CMSR reference.

6.9 Public Retrieval

One goal of the ODISS research plan was to evaluate how the general public could conduct their own CMSR searches on a digital image system. In the current system, the public is expected to use NARA's microfilm in the Microfilm Reading Room for genealogical research with only a minimum of staff assistance. In ODISS, it was hoped that the public could similarly do genealogical research by using a public workstation to conduct searches for Tennessee Confederate CMSR files, look at the digital images of the files on the screen, and print copies of the images on a laser printer next to the terminal.

To make the ODISS public workstation a self service activity comparable to the current use of microfilm, the technical requirements for the public station specified the following:

> "User instructions must be complete, logically written, and easy to understand by anyone without previous computer knowledge or experience ... the public's searches of the CMSR index must be guided and facilitated entirely by clear, simple, and courteous instructions displayed on the screen. These instructions must be particularly 'user friendly' since most of the public users, many of whom are elderly, have little knowledge of the CMSR records as well as no experience with computers and automated indexes. Public use of the system may be facilitated through the use

of menus, prompts, HELP screens, light pens, touch screens, mouse devices, or other easy-to-understand and use aids."

An ODISS public workstation was installed in the Microfilm Reading Room to collect data on the public's ability to learn and use a self-teaching reference station. However, the public workstation was never put into general service because it was quickly apparent that the general public could not learn to use the workstation in a reasonable amount of time. This was not the result of a deficiency in the basic retrieval subsystem but was primarily due to the on-screen instructions which could not clearly and easily lead the public through a self-teaching session. Consequently, data from a significant cross section of self-taught members of the general public was not obtainable. Other approaches, explained below, had to be devised to obtain useful data about the public workstation.

6.9.1 Test Procedures

One approach was to obtain detailed information on the reactions to the public workstation of people with some degree of computer literacy. These volunteers sat at a public station and taught themselves as much of the station's use as they could by following the on-screen instructions and employing their previous computer knowledge. As they worked, an NSZ staff member sat next to them to observe their learning process. They were encouraged to offer a running commentary on the workstation's virtues and defects as they explored the system. The NSZ staff member took detailed notes on these comments and on his observations of their use of the workstation. After the volunteers finished using the station, they were asked to fill out questionnaires on the ease of use of the public workstation (see Figure E-7 in Appendix E). Separate sessions were held with each volunteer, and the time of these sessions averaged approximately two hours.

The three primary volunteers for these tests each had some previous computer experience. Each had used word processing programs at work. One also had been the administrator of an on-line information system, who in that job had to think about the user interface to the system. Another person had used a spreadsheet program at work and had an Apple II with instructional and entertainment software for children at home. The third volunteer had only, in his words, "rudimentary PC exposure" other than using word processing.

6.9.2 Test Results

The volunteers' evaluations of the public workstation index and menu instructions were mixed. When asked to rate how well the public station worked overall on a scale of 1 to 10 with 1 being the lowest rating and 10 the highest, two gave the public workstation the rating of 7 and the other person rated the station as a 4. One person appreciated the fact that you could not make a "fatal mistake." If you pressed the wrong key, you could recover without a disaster - in contrast to the word processing program in his office where some errors cause havoc.

The more positive aspects of the workstation included the image quality and image manipulation features. When asked if the writing and printing on CMSR images were easy to read, one said reading the images was always easy and the other two replied that reading the images was usually easy. All felt the image zoom and rotation capabilities were easy to learn and use. However, two noted that there were no directions to explain how to get out of the zoom mode or how to scroll the enlarged image. One volunteer was impressed with the print function, saying the printer did "a really nice job."

While the volunteers were able to figure out various functions for themselves, they all needed some coaching from the NSZ observer at different points. Still, when asked to rate how easy or hard it was to learn the public workstation on a 1 to 10 scale with 1 = easiest and 10 = hardest, one rated it a 3, one graded it as a 4, and the third rated it as a 5. When asked to pick a verbal description of how hard it was to learn how to operate the station, two picked "average" and one chose "somewhat easy."

These relatively favorable ratings reflect the volunteers' own abilities to draw on their past computer experience to learn by trial and error rather than the clarity of the on-screen instructions. As the person who gave the station the most favorable rating for ease of learning wrote, "Introductory instructions are not clear - not particularly user friendly. For example, F1 or Help key is of no help." Another volunteer said the first screen of instructions was "daunting" and that viewers had to "wade through" a lot of verbiage. The most experienced computer user among the volunteers said that what she did in exploring the workstation was based on intuition grounded in her past experience rather than the directions on the screen. She thought that when a member of the public with no computer experience looked at the first screen of instructions, they would find it "a bit overwhelming" and say, "Oh my God, forget this."

The volunteers noted many specific deficiencies in the on-screen instructions. For example, movement out of the zoom mode and the significance of the original resolution mode are not made clear. There was no explanation of the CMSR records in the ODISS system, e.g., the war, time period, geographic area, or the kinds of information they document.

How to construct a search is not explained well. It was not apparent to one volunteer that a search can be made on a single index field, such as the last name. The use of NMI for no middle initial was not clear to another volunteer. All found the directions about using the numeric code tables for ranks, regiments, and companies incomprehensible; as one volunteer summed up the instructions on the code tables, they leave a person "totally baffled." Nothing tells the researcher about the importance of the cross references in the Remarks field; these should lead researchers to larger, more informative files on the soldier, but this is not mentioned in the instructions. Some of the weaknesses in the instructions were due to the fact that they were written in computer jargon rather than plain, simple English.

Another major area of deficiency was the failure to explain the use of the keys needed to perform many functions of the public station. The use of the function keys, labeled with the letter F (e.g., F1, F2) and positioned on the left side of the keyboard, was fairly easy to understand. However, the fact that the action or purpose of some of the function keys changed at different stages or menus within the public station caused some confusion and there were no instructions telling researchers about the changes.

A more serious problem was the failure to explain the use of the keys to move from one index field to the next for entering names, ranks, etc., to begin a search. The term "cursor keys" appears in the instructions, but this term is meaningless to people with no computer experience when it is not defined. There was also no explanation of how to erase or change an index entry, such as a misspelling of a name or an accidental entry of the wrong number for a rank or regiment.

The term FCN appears on most of the menus, but the fact that it stands for FILE CONTROL NUMBER is never explained. This may not be a serious problem for most operations, but it does have some impact during one method of printing files. If you select the print option

from the menu displaying a hit list of files, then you are required to type the file control number for the file you want to print. Without any previous explanation of FCN, this can cause confusion.

Besides the many deficiencies of the on-screen instructions, the volunteers noticed some other weaknesses in the public station. Two people felt moving from file to file through a long hit list was too slow, and one person suggested adding a capability to go directly to a file on the hit list. The lack of any apparent order in a long hit list bothered one volunteer. One person also suggested that a non-glare screen would be better than the present ODISS screens.

6.9.3 Analysis of the Test Data

There is a paradox in the results of the in-depth interviews. The volunteers voiced very strong and appropriate criticisms of the on-screen instructions of the public workstation. Yet they rated the station as fairly easy to learn. This is because their previous computer experience and coaching by the NSZ observer when at an impasse allowed them to progress despite the shortcomings of the on-screen instructions.

The retrieval functionality of the public workstation is essentially the same as the staff workstation. The menus, the use of function keys, the movement between file lists and images, the zoom and rotation capabilities, and the printing operation are all duplicated at both stations. As reported in the section on the staff workstation, archives technicians were able to make significant progress in learning the staff retrieval functions in a short period of time when given adequate training.

In other words, the actions and functions of the public workstation do not appear to be extraordinarily difficult. Good on-screen instructions with the current menus might make the station usable for the first time user. The volunteers for the in-depth sessions made some general suggestions for simplifying or redesigning the public user interface. One recommended that it be made more like the directions in programs for his children. Another suggested a series of simpler menus with fewer choices at each and with prompts along the bottom of the screen for each menu.

The assumption behind these suggestions was that the public station's basic activities of searching an automated CMSR index, viewing the images on a display screen and printing copies of the files are tasks that could be done by most anyone, if given the proper directions by the system. Any further work by NARA in the use of optical digital image systems should include a simple user interface that leans more heavily toward self-teaching.

6.9.4 Public Survey

To get a sense of public interest in using ODISS for reference work NSZ staff conducted demonstrations of the public reference terminal in the Microfilm Reading Room. These demonstrations used the Tennessee CMSR cavalry records that had been converted to optical disk storage. Indexes searches, files retrievals, image rotation, image zoom, and the quality of printed hard copies of the files were all demonstrated. Members of the public were asked questions about their opinions of the image legibility on the screen, the quality of printed copies, and whether they would want to use an optical disk system like the one being demonstrated. NSZ staff members or the people themselves filled out survey forms to record their reactions to the demonstration. Survey forms were completed for each person observing the demonstration.

In most cases the people simply watched the demonstration conducted by the NSZ staff. However, in a few cases people asked to try out the public station themselves, and these people were allowed to conduct searches and retrieve files with some coaching as needed by the NSZ staff members.

The reactions to the demonstrations were virtually all favorable and very often enthusiastic. For example, Marie Bigrau said that, "... even though I am 70 years old, I'm not too old to learn a new way of doing things .... I like it .... When you put more on the system, please add the state of Georgia." A professional genealogist, Edith Axelson, said, "... were that I was 40 years younger so that I could really make use of it ... marvelous!" She considered the images on the screen very easy to read.

Professional genealogists were especially interested in ODISS. The past president of the National Genealogical Society, Phyllis Johnson, thought ODISS is a "wonderful concept .... I hope you can develop it further." She liked the quality of the "beautiful and clear" prints. Marion Beasley, librarian of the National Genealogical Society, found the optical disk retrieval to be much faster than microfilm, adding that just before seeing the demonstration, "I've worn out my hand going through Arkansas microfilm." Looking at an image on the terminal screen, Mr. Beasley said, "That's beautiful." and, looking at a printed copy, he said, "It is as clear as it can be." Mary McCampbell Bell, a board member of a genealogical society, said, "We hope you get more funding to continue the project."

All the people liked the screen display quality for images. Several commented on the zoom capability. For instance, Marguerite Isman thought the images on the screen looked good and especially good with the use of the zoom. Charles H. Bibbing found the zoom particularly advantageous for magnifying signatures to make them more legible. Mr. Bibbing thought the optical disk retrieval terminal is very superior to the microfilm system in the Microfilm Reading Room.

All were impressed with the quality of the prints from ODISS. Many people commented on the fine quality of the copies of documents made on the ODISS printers. Lilla Licht, a professional genealogist, considered the prints to be better than the ones she generally could get from the current microfilm. William Carley called the prints "excellent." Mary McCampbell Bell, another professional genealogist, pointed out that the prints from microfilm on the reader printers in the Microfilm Reading Room are inferior to the prints from the ODISS printers.

When asked to compare the optical disk system to the current microfilm system in the Microfilm Reading Room, nearly all preferred a well developed optical disk system. Steven Green believed an optical system could be very useful for all genealogical research because of its speed and the legibility of the copies. Lilla Licht, Mary Neale, and Paul Verduin each found the searches to be much faster on the ODISS terminal than with the current microfilm. Barbara Heard generalized that using a computer is much quicker and usually easier, and that a computer-based system is more efficient than a microfilm reader. Judy Zieg recommended putting ODISS on a modem for easier access from sites outside NARA. Norma McGransee preferred the optical disk technology because it could improve the image of the original record and because it would save research time. And Janet Gant thought the optical disk system was "much, much better than microfilm in every possible way."

Only one person was clearly hesitant about implementing an optical disk system. While Sandra Napier believed it "looks promising," she felt an optical disk replacement of microfilm

for the general public should depend on how easy it would be to use by people not familiar with operating computers.

Others, who were enthusiastic about the potential of optical disks as a replacement for microfilm, also pointed to the need to make the public terminal user-friendly. Doris Boyd found ODISS much faster than microfilm and would prefer such a system, but had some qualms about learning to use the terminal. Both Ms. Boyd and Phyllis Johnson thought the screen needed to be put at a better viewing angle for comfort and especially for people with bifocals. Pat Bausell thought the system was great and gave optical disks her "vote," but still thought the public station should be made easier to use for "computer phobic" people.

Robert McKinney also preferred an optical disk system to microfilm, but suggested having both. He said, "We like the microfilm for backup and for those who may be afraid of computers; but this system is very good! I have used the Utah computer system for research and that is good also."

6.9.5 Analysis of the Survey Data

All the people who saw the demonstrations of the ODISS public workstation liked the optical disk system. They were favorably impressed by the image display quality on the screen and the high caliber of the printed copies from the system. People appreciated the zoom and rotation features for better viewing of images on the terminal's screen. Several commented that the ODISS prints were much superior to the prints from the microfilm reader printers in the Microfilm Reading Room. There were many enthusiastic comments about how much faster it was to conduct a search on ODISS than with the microfilm.

So, nearly all would welcome the replacement of the current microfilm system in the Microfilm Reading Room with an optical disk system having the capabilities they saw in the demonstrations. The only doubts about replacing microfilm with an optical disk system centered on the user friendliness of the public workstation.

The results of the public survey dovetailed with the results of the in-depth staff interviews. In both tests, the participants were glad to have the search, retrieval, and printing capabilities of the ODISS public workstation, but also wanted the learning of the system to be easier. The data from the public workstation survey suggests the conclusion that a public station for an optical disk system which is easy to learn and use would be preferred by a large and enthusiastic majority of the public to NARA's existing microfilm service in the Microfilm Reading Room.

6.10 System Manager

There were no tests designed for collecting data on the operation of the System Manager. Instead, analysis of this area is based on the experience of the users of this component. These users included the NSZ project staff, but the primary daily user throughout the test period was the NN employee designated as system manager when ODISS was installed.

6.10.1 Ease of Use

Operation of the System Manager component requires the use of three terminals. While the simplest activities at each terminal were easy to learn, each also had more advanced functions that required much more effort and aptitude to master.

System Manager Terminal

The System Manager terminal was used for most coordination and update activities in ODISS. Database operations performed at this terminal included correcting errors in the index records of CMSR files. The terminal was used to control the system's functional modularity by changing the task assignments of workstations, adding new users to the system, and granting or limiting the users' access to different functions. Management reports on the performance of all major functions were obtained at this terminal. Another major use of the terminal was for resolving problems with blocks of files as they passed through the different input functions.

Most activities performed on the System Manager terminal were directed from easy-to-use menu options. Once the user was familiar with general system terminology, these options became largely self-explanatory. Grasping the functions of the add, delete, inquire, and modify modes were even less difficult for the average user. Commonly used key actions were provided on the bottom of all System Manager Main Menu screens, a convenient reference for the novice user.

Some activities were not run from these menus. Instead, the system manger left the menu-based module and entered the UNIX shell, where short commands were used to execute programs performing various tasks. These shell commands and programs were developed by the contractor for certain activities not covered in the basic menus. Most of these shell commands were used with sufficient frequency that the user quickly memorized them. Performing operations on the shell screen was very easy, requiring only key-entering and executing the desired command.

During the test period it proved useful, and even necessary in certain instances, to develop new programs to make database inquiries or perform other tasks for which the contractor had not provided. To accomplish this, an NSZ staff member had to learn at least the rudiments of the UNIFY database management system's version of the Standard Query Language (SQL) as well as some UNIX commands and the basics of the UNIX line editor (ED) and screen editor (VI). This was largely a self-teaching exercise. The resulting programs were made available to the designated system manager and the few other users of the system manager component. Because these programs could be left on the screen by a previous user, or entered accidentally by inadvertent invocation of relevant commands, it became necessary for all System Manager users to acquaint themselves with the proper "exit" or "quit" commands.

CSE/ARS Terminal

The CSE/ARS terminal was used to track system activity and file storage conditions on both the temporary magnetic Capture Storage Element (CSE) and the permanent optical disk Archive Storage (ARS). Most CSE or ARS tasks were relatively simple and may be done with alacrity once the user possessed minimal knowledge of the functions he wished to perform. However, maximum use of this terminal, especially in its CSE capacity, required a detailed knowledge of the numerous abbreviations employed as commands for various activities. For examination of the Dump Sectors selections on both CSE and ARS, conversion from hexadecimal to decimal numbers was unavoidable. Therefore, users had to be capable of using a hexadecimal calculator.

Initiation and Monitor (IMS)/Archive Control Terminal

The IMS terminal was used for initializing the UNIX operating system and starting the core computers in booting[72] ODISS. It also handled various transactions in the management of data movement between the optical disks and the other components of ODISS such as the temporary magnetic storage and the workstations. This terminal was also employed to implement commands for nine operations performed with the optical disks. These included mounting or dismounting (i.e., removing) optical disks from the jukebox or the stand-alone optical drives. It was used to format or initialize new optical disks and to make backup copies of optical disks. Writing directories to the disks and diagnosing the optical media were two more actions performed at the IMS terminal.

For the nine standard optical disk operations the IMF terminal was extremely easy to use. All nine commands were listed in the ODISS Operations Manual. Any user with cognizance of only the most minimal ODISS features could enter the proper parameters in response to the prompts on any IMS screen. System manager personnel had little difficulty learning to use IMS options.

The ODISS initialization and system boot activities were more complicated. Users were able to initiate the core reset unassisted, after only a few performances of that operation. The full core reboot was more complex; but by following the procedures outlined in the ODISS Operations Manual, new users became adept at its execution in a short time.

Other actions performed on the Archive Control terminal required executing programs developed by the contractor. As long as clear instructions were provided, the designated system manager was able to run them without difficulty. Such was the case with the disk diagnosis and the procedures to recover from that program.

6.10.2 Specific Problems and Suggestions for Improvement

Although the System Manager and its three terminals generally performed well, some problems did exist. Correction of these problems would simplify some functions and increase overall ODISS efficiency.

A significant problem was the continuing inaccuracy of certain management accounting reports. Discrepancies were found in the category totals for different reports on the same workstation for the same range of dates. Corrections to the programs eliminated this problem in the daily and weekly reports, and all but one of the effected quarterly reports were corrected. Because the modified program for the yearly report still does not execute correctly, modification of that program will be necessary for the production of more accurate annual summaries.

A serious shortcoming of the System Manager was the lack of data on optical disk addresses for non-CMSR files. Such data is available through the ARS function of the CSE/ARS terminal, but only after a disk directory has been created upon the completion of archiving on a disk side. Without readily retrievable sector address data, examination of an archive sector, a requirement for recovery from an abnormal archive termination, is considerably

[72] *Booting* refers to the process of initializing computer operations once electric power has been applied to the system equipment.

more difficult. An optical sector examination for a non-CMSR file can now be performed if the sector for the last prior CMSR file on the disk side can be located. Beginning with the CMSR file, the pointers can then be followed from sector to sector until the desired file or the end of archive data is reached. If such a search is necessary, it is presently advisable to keep a written record of non-CMSR sector addresses for future reference. This search process can be extremely time-consuming and is made even more difficult if no CMSR file is available for use as a starting point. If possible a means of locating the sector address for each non-CMSR file immediately after archival should be developed.

Another area where improvement could be made is the CMSR screens. At present no screen is available that displays written information for numerically coded index data together with the remarks field and optical disk information. Because it has the latter two features, CMSR5[73] is generally used both for index verifications and corrections, and for study of archive failures. Since CMSR5 does not show the written equivalents of numeric codes, indexing errors for numerically coded data, especially in regiments where the company codes do not follow the standard coding order, have an increased chance of going undetected. Development of a single screen featuring all this data would be helpful in catching numeric code errors.

Lack of a backspace capability, which does not also erase all previously written text through which the cursor is backspaced, is a minor problem on the System Manager terminal. A similar inconvenience occurs in the CMSR5 mode, which is used to correct errors in the index records for CMSR files; text in the remarks field of the CMSR5 screen cannot be revised without erasing and retyping the entire field. Provision of a cursor which can be moved through the remarks field of the CMSR5 screen and backspaced through the other screens without erasing good text would avoid the need for reentry of previously acceptable text, reduce the potential for new errors, and provide for quicker and easier corrections.

The Archive Control keyboard also shares, with the System Manager keyboard, the inability to backspace without erasing all text back to the point where the cursor is stopped. But, because responses for IMS prompts are so brief with no present response requiring more than five characters, absence of this feature has been only the most minor inconvenience at the Archive Control terminal.

The present sequence of block processing stages: entry, index, quality control, pre-rescan, rescan, pre-archive, and archive, has proved quite satisfactory. The occurrences of incorrect block and/or file stage assignment by the system were largely eliminated by modifications made during on-site testing. The few instances of this problem occurring since that time were easily solvable by trained System Manager operators. Otherwise, the only problem arising with regard to the processing sequence was the occasional need to return a block to a previous stage. This generally arose either from operator error or from damage to digital data sustained during a terminal or system malfunction. Addition of the pre-rescan and pre-archive programs decreased the need to return blocks manually to previous stages at the System Manager terminal and increased early detection and correction of stage-related problems.

[73] Refer to page 271 for additional information regarding CMSR5.

6.10.3 Overall Effectiveness

The System Manager functionality was a combination of ease and inadequacy. As described above many of the basic operations were simple, menu-driven actions that were easy to learn and to perform. But many other operations were substantially more complicated and required learning a lot of complex information rapidly for the successful operation of the system. The project was lucky that the designated system manager, who had virtually no computer background, was a quick learner who could work effectively under considerable pressure during the early months of installation and debugging.

The System Manager showed some signs of incompleteness. Some of these weaknesses are still evident and were discussed in section 6.10.2 on page 146. Others were corrected during the months of shakedown and debugging by the contractor immediately after installation. Some other deficiencies were solved in-house by developing procedures to correct certain problems in the database records and to get important information on certain aspects of the system's performance not covered by the standard management reports.

Many aspects of the System Manager worked well. The major lesson from the deficiencies in the ODISS System Manager is that more capabilities should be provided and they should be accessible in simpler formats. Otherwise, successful coordination of the system will depend too much on having the good luck to find a good operator that can master undue complexities in operations and compensate for unaccountable gaps in functionality.

6.11 Remote Workstation

As part of ODISS, a remote workstation was installed in the Tennessee State Library and Archives in Nashville, Tennessee. Since the Tennessee CMSR records were scanned, Nashville was considered an ideal location to test remote index search and retrieval. Due to cost factors, the digital images of the scanned paper files were not accessible in Nashville. However, the Tennessee Archives had a complete microfilm copy of the Tennessee CMSR holdings. The purpose of placing an ODISS terminal in Nashville was to test the regional use of a hybrid retrieval system consisting of remote access to an automated index coupled with local access to microfilmed genealogical records directly relevant to the immediate geographical area.

6.11.1 Workstation Configuration

To accommodate the requirement for a remote terminal index search capability, a Unisys PC and printer were connected to the System Manager Subsystem via telephone lines and modems. This terminal could search the ODISS index database and print the results on a dot matrix printer. With that information, the user could go directly to the corresponding microfilmed file. The ODISS design specified that, if the user desired, hardcopy prints of requested images could be created in the ODISS facility and forwarded to Nashville.

6.11.2 Operational Experiences

The remote terminal was delivered directly to the Tennessee State Library and Archives. Installation and operational instructions were provided with the equipment, as part of a package designed for remote site clients. This user-installed concept was successful, when augmented with further verbal and written guidance. In initial efforts to access the ODISS

system, the remote terminal experienced problems. These problems were addressed by Unisys engineers while on-site at NARA.

NARA provided the workstation to gain information about the ease of learning the index search procedures, workstation ease of use, clarity of screen instructions, and usefulness of index data retrievals. Knowledge about workstation operations by genealogists unfamiliar with computers and keyboards would prove useful in planning future system configurations.

The Nashville Archives staff were successful in accessing the full search capabilities of the system They accessed the code tables, and printed search information using the dot matrix printer. The staff reported that the index-only information was useful as the workstation was installed in their microfilm reference room and provided another source of determining information about the CMSR files. They expressed a desire to obtain access to the digital images directly at the remote system site if possible sometime in the future. Workstation utilization was limited since only the Cavalry records were converted and because of system problems relating to log-on to the main ODISS system. Resolution of system problems was difficult because of the physical distance from ODISS equipment maintenance support. Communication line options are other considerations which would need to be addressed in any planning of an expanded system configuration. Nevertheless, the Nashville staff remains interested in continuing with testing and usage of the workstation.

6.12 Production and Evaluation of Image Quality

A user's ability to retrieve accurate and understandable data is the main objective of an information processing system. Whether data is stored in paper, microform, magnetic storage, optical disks, or any other format, the ability to identify, locate, and interpret this information is critical in determining a system's effectiveness. Microforms have historically served as the main alternative to paper for image storage within the National Archives.

One of the main reasons for acquiring ODISS was to determine how well digital imaging technology handles NARA's varied document holdings. It is inevitable that any new technology under consideration to replace or augment existing applications would be required to demonstrate clearly its capabilities. Since images are stored both in NARA's existing microform and the ODISS optical disk system, then a comparison of image qualities is important for determining system performance and how well these systems meet NARA's mission needs.

Correlative analysis of images produced from different technologies, such as comparing microfilm images displayed on film viewers to optical disk images on high resolution video display terminals, is not a simple task. There are many subjective factors which the observer brings to and is confronted with during an imaging evaluation. Factors which influence microfilm quality include (but are not limited to) raw film stock resolving power, optical and mechanical performance of the camera system, overall film quality, film processing conditions, film generation and polarity, microform viewer optical qualities, and viewer-printer hardcopy production. Electronic imaging elements which influence image quality include scanner CCD dpi resolution, threshold settings, dynamic/constant threshold selection, CCD spectral sensitivity, video display screen resolution, and hardcopy output capabilities. These criteria should be considered when comparisons are required.

6.12.1 Technical and Subjective Considerations

Creating quality images requires that attention be directed to production criteria. The best images are produced in those systems with quality assurance programs, employing at least a three level approach:

- Verifying system performance
- Monitoring quality of the integrated system
- Inspecting images to verify conformance to established parameters, guidelines, or standards

System performance relates to specifications and the performance of the system as delivered. For example, ODISS procurement documents specified NARA's performance expectations. Factory and on-site validation by system engineers tuned each ODISS component to effect optimum performance, using specially designed calibration tools. These efforts helped produce the best output quality available.

The second element is process control monitoring of the total system performance. Maintaining consistent quality is difficult unless the complete process is controlled under stringent specifications. A separate quality assurance monitoring staff, reporting to the system manager, might be required in a larger system. This staff would monitor system performance and product quality, and be responsible for image inspection and rejection rates, equipment operation and calibrations, and monitoring operations staff performance.

The third factor is ongoing product inspections, which can vary from checking every image to random sampling or perhaps no inspection at all. It is preferable to identify a problem immediately to ensure timely corrective action. ODISS quality control operators visually compared each original document to its corresponding video screen image. ODISS guidelines specified that document images must be readable at the 150 dpi level and should be rejected if zooming was required. Operators were also checking for missing images and other problems.

A related issue is the establishment of image quality guidelines. If established quality expectations are too high, then production throughput may be degraded in an effort to meet these goals. On the other hand, too low of a quality baseline could lead to poor quality or non-readable images. Once established, this quality level must be implemented through the production staff. It is of little value if operations staff are unsure about exact station requirements. Marginal documents are the most difficult to capture faithfully, and therefore usually require more visual analysis. Understanding image acceptability is vital to maintaining throughput while still producing quality images. Management should routinely monitor system operations to verify adherence to standards. This is especially important when operators are rotated between stations.

6.12.2 Image Enhancement Issues

It is important to consider image "enhancement" within the confines and requirements of the individual application. When a document is scanned, it is modified based on the threshold level and other criteria selected by the equipment operator. In an ideal situation all the information is captured, and the extraneous information is discarded. This would "clean up"

documents with stains and other defects and increase legibility. The scanner's electronic thresholds compare pixel reflectance values to a selected set of neighboring pixels. This process determines whether each individual pixel is a "one" or a "zero." A debate may ensue around just what constitutes "information" and who should decide what is "extraneous" information or non-data.

One of the generally accepted benefits of digital imaging technology is the ability to clean up poor quality original documents. In the case of a severely stained document, most would argue that removing the stain is beneficial to the information legibility efforts. There may be some, however, who would argue that the stain also constitutes information and should not be eliminated. It is not clear what happens to this argument when a document is faded to such an extent that textual information is no longer readable. Would some insist that the image be left in a likewise unreadable condition? Or would they, as an alternative, permit the power of digital image technology to make that image readable and therefore useful? Perhaps the passage of time, and increased awareness of the digital imaging capabilities, will allow acceptance of the concept that an exact facsimile of a document may not always be desirable.

A decision as to what a system will capture should be based on testing with the original documents. If all documents are recent, office-type originals, then this is probably not a major issue. NARA's holdings, which in many cases contain stained, faded, and deteriorating documents, make system tuning factors more critical. The question of how close the digital image must come to the original (with all its faults) must be decided by the professional archivist. The results of the ODISS testing suggest, however, that digital imaging technology can offer an excellent and fully readable image that can be substituted for the original in all but the most recondite and atypical situations where the copy must be as close as possible to an exact duplicate of the original document.

6.12.3 Photographic and Electronic Imaging

Microfilm manufacturer's guidelines are based on typical modern office documents with special technical adjustments being required only for documents with atypical attributes. Just as film emulsions have slight sensitivity variations, CCD electronic elements differ slightly in performance characteristics. Film products may vary, but usually fall within tolerances set by factory production quality control guidelines. The most consistent film materials result from the manufacturer's ability to control the entire production process. Modern microfilms are high quality products manufactured to meet typical users requirements.

In order to improve image capture, optical filtering is one possible approach. This technique has been successfully used in photographic copying to remove image stains or other defects. NARA has used this technique to improve image qualities when microfilming special holdings. ODISS testing showed that a lens filter could improve an electronic scanner's ability to capture colors. These special techniques usually require some trade-off in production rates versus image quality. If large numbers of records are exactly the same in image characteristics, then the input device can be optimized for those records and rapid conversion throughput is possible. If the records vary widely in appearance, they may require individual settings which can greatly slow down the conversion process. ODISS production utilized green optical filtering for the entire conversion effort. No time was lost in changing filters or performing other mechanical optimizing procedures. The operators used only pushbutton-activated electronic controls based upon document appearance.

6.13 General Testing Issues and Results

6.13.1 Validity of the Original Design Concept

The basic design structure of ODISS was based on sequential operations. The effective employment of the system was dependent upon the successful completion of the preceding step. That is, in order for the quality control stage to perform its function, the file and block must have been indexed. Additionally, to be able to input the index, the file and block must have been scanned. The logic of the system flow began with the wholesale scanning of all the file pages within the block of work. This step utilized the high speed scanner to gain the optimum conversion of documents per hour.

The tabletop scanner could be used as an original entry device, as well. In applications that required a great deal of special handling for the entry documents, using a number of the platen-type of tabletop scanner placed in parallel may offer a suitable input option. However, for input documents that can be scanned using high speed paper transports, one or more high speed paper scanners may be appropriate for the application.

The scan density chosen for the high speed scanner was a moderate, 200 dpi since this density creates smaller image file sizes than higher scan densities. This resolution has traditionally been used for office-type documents with the higher resolutions reserved for those applications requiring more detailed image definition. This density turned out to be more than adequate yielding a low image rejection rate (due to poor quality) of less than five percent.

The ODISS design located the index and quality control operations after the document was scanned and the image created. This approach is the most common for those applications that do not already have an existing computerized index that can be referenced and interfaced with the new image files. The method of scanning the documents and then using the image as the source from which the index is created, provided the system with a smooth and orderly transition from the paper file to the electronic image file.

Recovery of poor quality images is an important consideration in any digital image based system. In the ODISS design, it was recognized that there would be a certain percentage of images that would be rejected on the basis of inadequate quality from the initial scan conversion. The choice was made to use a separate scanner for rescan, instead of merging the rescan documents back into the production stream, in order to be scanned again by the high speed scanner. By using a separate scanner for rescans, it was possible to continue to gain the maximum throughput possible with the primary (high speed) scanner and only use extra finesse on the problem document images.

The general ODISS design as defined by the functional requirements was quite efficient. The contractor, however, chose to implement the design in a scheme that did not prove to be as efficient as it should have been as based on the original design requirement. One of the most obvious examples is based on the integration of the capture and retrieval subsystems into one control unit. This served to slow down both input and output operations. This may have been better conceived as completely separate entities. The input conversion system would work independently of the output retrieval system. The drawback to this design strategy is some delay in allowing access to the newly converted files until they have completed the entire conversion process that includes indexing, quality control, and rescan.

6.13.2 Modifications to System Operations and Workflow

As experience was gained running ODISS, some aspects of the system and its operations were changed. The modifications were made to experiment with different techniques and to improve the system's performance.

Perhaps the most important modification to ODISS was the addition to the low speed paper scanner of new image capture and enhancement hardware and software from Image Processing Technologies (IPT). The IPT firmware and software were superior to the image enhancement algorithms provided by Unisys. Use of the new capability resulted in better image quality for difficult documents than had been achieved earlier. There were also some savings in storage as, in a number of cases, the enhanced, more legible image from the upgraded low speed scanner took up fewer kilobytes of disk space than the original image from the high speed scanner. As operators gained facility with the new capabilities, speed of processing improved at the low speed scanner since fewer retries were necessary.

Another significant modification of the ODISS input processing was the addition of the pre-rescan/pre-archive program written by Unisys. Reboots of scanners and workstations could cause the system to record inaccurate information about the number of pages in a file. These inaccuracies in turn could cause problems with retrieving files at the low speed scanner with the results that some pages needing rescanning were sometimes missed and that some files whose pages should have been rescanned were archived, i.e., written to optical disk without rescanning bad images. The new program was run on each block of files at two points, between the quality control and rescan operations and between the rescan and archiving operations. The effect of the program was to catch files that should go to the rescan station for image quality improvements and to make sure that files were not archived if they still contained images that needed rescanning.

There were several changes at the high speed scanner to cope with different problems that arose during the processing of the CMSR records. The most significant at the high speed scanner was the definition of the settings of the control panel's buttons to match the characteristics of the CMSR records. The scanner's buttons included a variety of image capture choices, including three possibilities for document size and nine different threshold selections each for the fronts and the backs of documents. After a review of the records, the size buttons were set to match the three most commonly occurring sizes in the CMSR files. Analysis of a large number of CMSR files resulted in the development of a list of the optimum threshold choices for documents colors and contrast levels most frequently occurring in the Tennessee files. This list was posted at the scanner to help operators quickly select the best image enhancement setting as documents arrived at the station.

More minor changes at the high speed scanner involved operational procedures. A good example of these operational procedures concerns documents not sent through the scanner. A few documents were not sent through the high speed scanner, either because they were too large or thought to be too fragile. When it appeared that the quality control operators might miss these skipped pages, a rescan sheet was developed and sent through the high speed scanner in their place. At the quality control workstations the operators then saw the sheets reminding them to mark the place in the file for adding the skipped documents at the rescan station.

Another change in system operations derived from the discovery that the lists of companies within regiments were sometimes incomplete. While indexing the newly entered files,

indexing operators found companies that were not on the code table lists of valid companies for that particular regiment. While it was easy to add the new companies to the regiment's company table at the system manager terminal, the initial procedure for downloading the revised code tables to each workstation was very time consuming. Downloading the tables at a workstation could take one hour or longer. This had the potential for totally disrupting operations if all the workstations were tied up with the original, lengthy downloading procedure. Consequently, a new procedure was developed to use new copying programs on a floppy disk.

Once the potential severity of the disruption to normal operations was recognized, the new programs were written by a Unisys programmer to ease the problem. The tables were downloaded to a single workstation, which remained a lengthy task. But then using the new programs, the revised tables were copied quickly to a floppy disk and then from the floppy disk to all the other workstations. This procedure required only one workstation be taken out of normal operations instead of all the stations, and it made the transfer of the corrected tables to all the other workstations into a quick and easy job.

6.13.3 System Modeling

As noted earlier, during the early stages of the production operations testing, it became obvious that certain inefficiencies in the ODISS system implementation were retarding system throughput. Operators at the various system workstations found themselves having to wait while the system completed certain actions before they were permitted to continue work. These periods of wait time were biasing the workload statistics with non-productive time. Therefore, it became necessary to seek outside assistance to capture accurate and unbiased data on system throughput. Besides compensating for the wait time problems, it was also necessary to consider that ODISS had certain elements within its design that would not be duplicated on a large production system. Therefore, it was equally important to determine how a future system configuration should probably appear and to gather data on that design approach.

In early 1989, NARA contracted with the Navy Regional Data Automation Center (NARDAC) for the services of a professional operations research analyst and statistician. In order to gather the most useful data possible on systemic operations and design configurations, the NARDAC analyst programmed and used a system model to simulate system modifications and design enhancements.

6.13.3.1 Operational Use of the Modeling Software

NARDAC's analyst developed software that would enable simulation tests to be conducted and timed for the effects of alternate configurations. Predetermined portions of the system would be modified within the constraints of the model in order to acquire more accurate test data. The system model became invaluable in identifying ODISS system bottlenecks and design deficiencies and eliminating their influence on the production statistics.

Since the modeling software consisted of a timer program and related configurational analyzers, testing with the model involved setting up various operational sequences and

timing the work[74] and wait[75] times. It was determined that the system suffered several intervals of inordinately lengthy wait times. One example of this wait time on the high speed scanning operation was the time from the point that one file was closed and the next file was opened (in order tc scan in a new CMSR file). During this period of time, the operator could not accomplish any additional work. Therefore, this wait time represented wasted time.

The ability to model the conversion subsystem enabled the research team to identify the work and wait times for each of the operations. Individual testing was carried out with similar files. Four separate sets of timings, as documented in Table 6-25 on page 157, were taken for each operation within the conversion subsystem.

6.13.3.2 Analysis and Findings

As was noted earlier in section 6.2.2.2 on page 92, in Unisys's implementation of ODISS, the length of time required for document capture is directly related to the number of page images present in the document files. As the number of images in the file increases, so does the capture speed per image. This phenomenon is due to the frequency of file openings and closings. As the number of images per file decreases, the corresponding increase in the frequency of file openings and closings places an additional burden on the system database. The result is extended workstation wait time while the database software modifies and closes the current file's database record, and then opens a new record and allocates temporary storage to receive the scanned images from the next document file. Conversely, if the files were to contain a greater number of page images, the result would be fewer database transactions and a commensurately reduced database load that could be handled by the system without long periods of wait time at the workstations.

The use of the computer model made it possible to determine the production throughput rates that would have been possible if the ODISS system had no delays at the operator workstations. As addressed in the next section, it also was able to evaluate the operators' actual work times at the various workstations in order to determine the best task configuration of workstations to maximize throughput.

Table 6-25 documents one of a series of process timings that were conducted and analyzed using the computer model. It shows timings for the processing of four blocks of CMSR files as they progressed through the high speed scanner, indexing, and quality control stations. System operations were fully loaded during the tests in that all system workstations were in use. Operators at the various workstations attentively performed their respective tasks during those segments of time that the system would permit them to do so. Table 6-25 also shows the analysis of the timings data using NARDAC's system model. The blocks of files used in this particular test were typical of the CMSR sample except that they averaged only 2.57 images per file.[76] The average ratios of work time to the non-productive wait time ranged from roughly 2 to 1 in the case of the high speed scanner to less than 1 to 1 for the

[74] The amount of time actually necessary for an operator to perform a specific operation.

[75] The interval between the periods when an operator is able to key-enter data or commands at his terminal. During the "wait" period, the system is performing certain actions that temporarily disable the terminal.

[76] As noted in section 6.2.2.2, the average number of images per CMSR file was determined to be approximately four.

in stations. The average time to complete the entire linear process (i.e., taking a complete file through all three task processes), was 108 seconds, of which 42.9% was wasted overhead (depicted in Figure 6-11). If the ODISS pilot system had been implemented with sufficient machine resources and concurrent task processing capability to eliminate all time delays at the workstations, then the total linear time to complete a file would have averaged just over a minute.

6.13.3.3 Optimum Workstation Configuration

The ODISS system was designed so that any of the PC-based workstations could be utilized for indexing, quality control, or retrieval functions. The number of workstations assigned to each activity could be modified from the System Manager station according to the need at the time. The data gathered from performing the modeling tests provided information for selecting the best component configuration for optimum throughput. (See Figure 6-12.)

Workflow Data from System Model

Number of Files in Block	Images Per Block	Work-Station	Work Time Per Image (sec)	Work Time Per File (sec)	Wait Time Per File (sec)	File Name
60	155	HS	6.91	17.86	5.42	HSWW1
59	151	HS	8.11	20.42	7.16	HSWW2
56	143	HS	4.86	11.58	8.15	HSWW3
58	151	HS	6.03	15.18	11.14	HSWW4
60	155	IND	9.21	23.78	24.77	INDWW1
59	151	IND	9.94	25.01	24.51	INDWW2
56	143	IND	8.86	21.11	25.53	INDWW3
58	151	IND	8.65	21.76	26.41	INDWW4
61	155	QC	10.98	28.36	12.23	QCWW1
59	151	QC	14.00	35.24	12.55	QCWW2
56	143	QC	5.38	12.83	13.12	QCWW3
58	151	QC	5.40	13.59	14.53	QCWW4

Composition of Sample

Average Files per Block:	58.2500
Average Images per Block:	150.0000
Average Images per File:	2.5751

Averages of Times at Workstations

HS:	6.48	16.26	7.97
IND:	9.16	22.92	25.31
QC:	8.94	22.51	13.11

Average Total Linear Times

Work Seconds per File:		61.68		
Wait Seconds per File:			46.38	
Total Seconds per File:				108.06
Work Seconds per Image:	23.99			
Wait Seconds per Image:				18.04
Total Seconds per Image:				42.03

Table 6-25

Total Operational Time per File

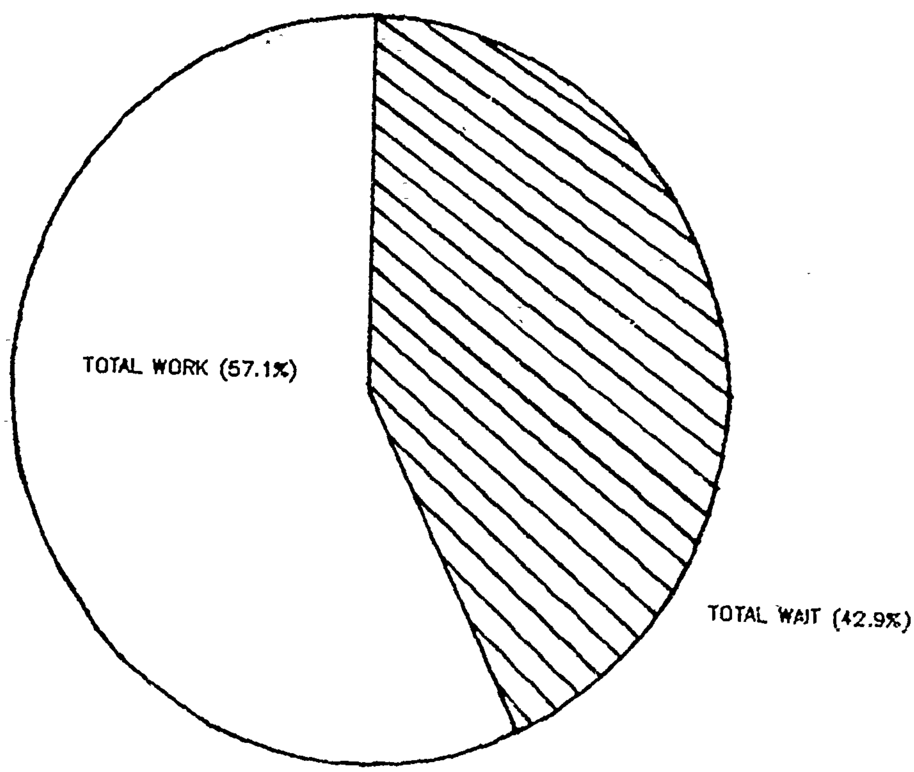


Figure 6-11

Optimal Workstation Configuration

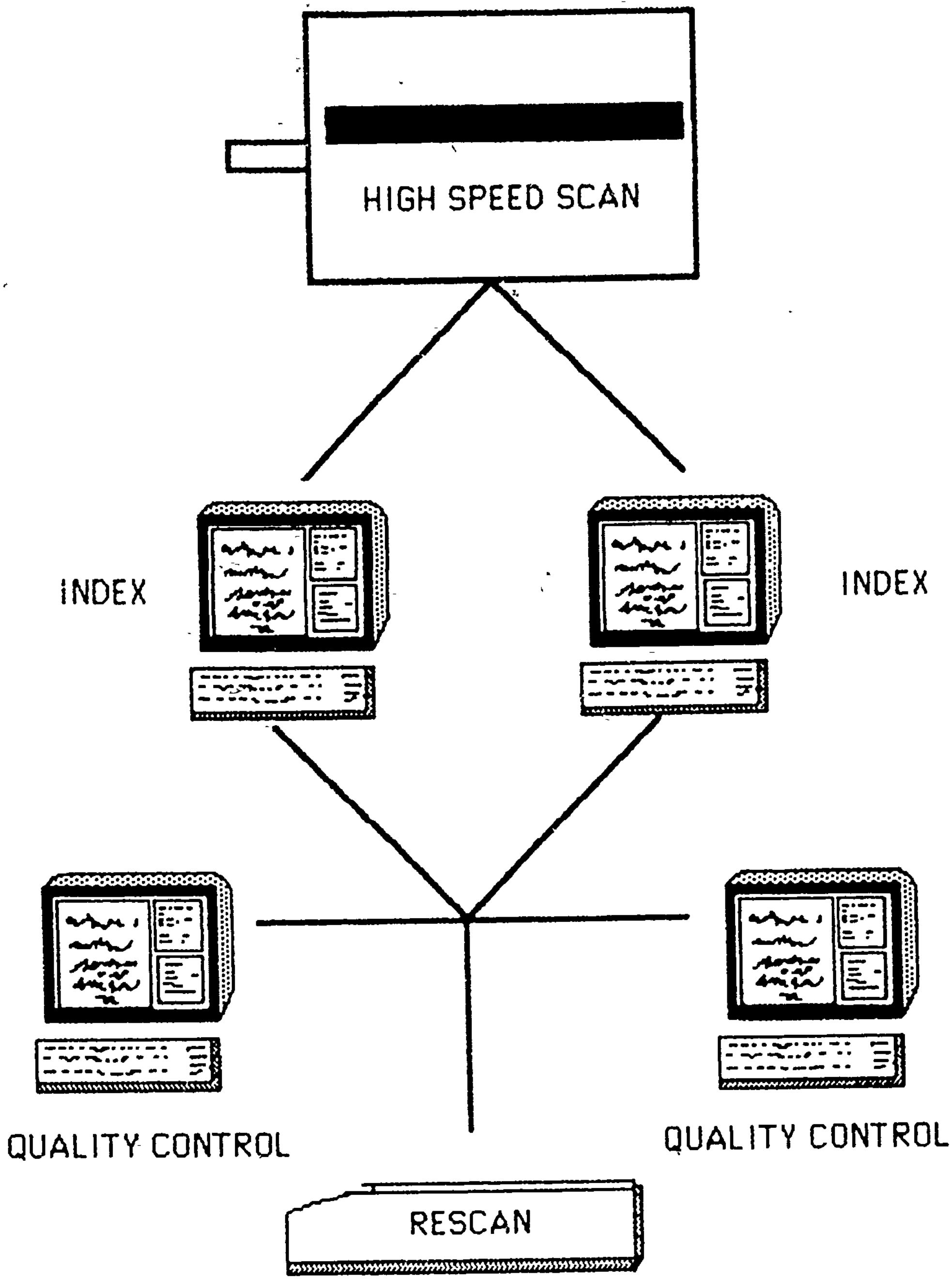


Figure 6-12

The averages in Table 6-25 were typical of those that were modeled. Note that the productive work time averages for the index and quality control stations are roughly the same and that each is somewhat less than twice the average for the high speed scanner. This means that the digital images created from one high speed scanner can be indexed by two index workstations without creating a backlog. The indexed files from those two workstations can be reviewed for quality with two quality control workstations. Therefore, the optimum configuration of the conversion subsystem, utilizing existing ODISS hardware and software, would be two index and two quality control workstations for the one high speed scanner. This holds true regardless of whether the wait time figures are factored in or out.

6.13.3.4 Optimum Performance Potential

The model also permits a projection of the maximum throughput that could be expected from the ODISS system with its current software inefficiencies, as well as subsequent to any improvements that achieved elimination of the time delays at the workstations. Although the number totals in Table 6-25 are for linear timings; in reality, all the workstations operate concurrently in parallel. Since in the optimum configuration consists of one high speed scanner and two each of the index and quality control workstations, the effective work time averages for the latter two stations would be halved (i.e., 11.46 and 11.26 seconds instead of 22.92 and 22.51 seconds per file respectively). When these figures are compared to the average for the one high speed scanner (i.e., 16.26 seconds), it becomes apparent that the high speed scanner is the potential bottleneck when all processes are conducted in parallel (or pipeline) fashion.

If a production workday consists of seven hours (or 25,200 seconds), then the figures from this application of the model suggest that the ODISS system is capable of processing 2633 images per seven-hour day with its current time delay problems, or 3890 images per day if the time delays were eliminated. In reality, this particular set of numbers represents only one application of the model and cannot be considered as statistically reliable. Nevertheless, it is interesting to note that on two different days during the ODISS test period, the operators were able to process in excess of 3000 images in a seven-hour period (and that was using the system replete with its inherent workstation time delays.)

6.13.4 System Maintenance

Maintaining a system, with the complexity of ODISS, required a great deal of planning and foresight. Options for system maintenance generally fall into categories of on-site versus on-call access. The original specifications for ODISS called for on-site maintenance for the first year and on-call maintenance (at NARA's option) for one additional year. The price bids included both of these costs as a part of the total system cost. This approach gave NARA the type of immediate response that was necessary for the initial period of shakedown during the first year. Also, the main part of the testing was scheduled for completion during the first year. It was imperative that the system be up and running for as much time as possible during that critical period.

The testing program for the second year did not necessitate the more expensive requirement for immediate response. Slower on-call reaction times were appropriate in order to conduct the remaining tests that were not production oriented and, therefore, not as time-critical. On-call with a minimum response time of two hours proved adequate.

6.13.5 Personnel and Staffing

Although the majority of this report discusses the hardware and software aspects of these technologies, personnel considerations are equally important. This section discusses some of the experiences with the operations staff doing the file conversion segment of the ODISS test.

6.13.5.1 Training and Operations

Training for the ODISS operations staff consisted of two separate and distinct types. One week was spent for formal training and system familiarization. This was followed by on-the-job assistance for several weeks. The more formalized training was adequate to acquaint the new staff with the technologies involved and the operation of the system's components. As the operators began working with the system, it proved to be very useful to have the system contractor and designers on-site and available for immediate assistance and instruction. As new employees rotated through the operations staff, they learned the component procedures by monitoring the work of the experienced members of the staff during normal daily operations. The combination of formal training followed up by informal on-the-job training served well as the ODISS training format.

6.13.5.2 Cross-training

ODISS was designed to allow flexibility in many of its operations. The workstations could be configured to work as index or quality control terminals depending on the needs of the system at that time. In order to utilize this flexibility in the operation of the system, the operators needed to be able to perform any of the functions of the system. As a result, the operators were trained on each component in the system. This cross-training provided the means to rotate the operators through the different system functions. Rotating the operators enabled them to maintain their recollections of and efficiencies with the usage of each system component. Since repetitive work can be boring, this rotation tended to provide the additional benefit of helping the operators focus their attentions on their assigned tasks without suffering too much boredom or fatigue.

As time passed, however, it was evident that certain operators had either an affinity for or an aversion to a particular component. Others did not display the degree of skill necessary on some of the more complex tasks. The frequency of rotation was slacked until several days would pass without switching positions. The idea of training all the operators on all aspects of the operation of a system is sound for backup or emergency situations. However, the talents and desires of the operator should be taken into account in order to allow them to work, when possible, on the component of their choosing.

6.13.5.3 Operator Performance

No matter how a system is designed, human performance plays a significant part in the measurement of systemic efficiency and operational throughput. Several important lessons were learned in the area of staff performance during the ODISS tests. The most important lesson involves the necessity of instituting guidelines and expectations for minimum levels for speed and quality of performance. These guidelines could be administered by the use of incentive programs. Without this procedure, it proved difficult to maintain an acceptable level of performance from the staff operators.

In addition, the ODISS test management was essentially divided into research and operational aspects. The day-to-day operations were administered by units belonging to the Office of the National Archives. At the same time, the project management and daily testing were managed by the Archival Research and Evaluation Staff. Because of this dual management approach, the operators were unclear as to who would administer the room rules and regulations. As a result of this confusion, abuses of work schedules and on-duty privileges were common.

One issue of possible significance involved the wait time at the workstation that appeared between files. If the operator used this twenty or thirty second period to read a newspaper or book, some time would be lost when the system was once again ready to accept entry. The operator would have to refocus his attention back on the operation. If workstation wait times were reduced or eliminated, the problem of operator attention span would disappear. At the same time, it is highly desirable that all extraneous activities be left outside the operations area.

6.13.6 Ergonomic Factors

During the production sequence of the ODISS testing, the operators in the room pointed out several areas that required improvement involving ergonomic factors. Some of the comments centered on the coolness of the computer room. The room was kept at a fairly constant 70 degrees. The problem was not one of temperature, but of air flow. Five air handlers were required for the room. These units distributed a moderate flow of air up the wall and across the ceiling. This method of air distribution created cool drafts (even on the lowest blower fan setting) that, while accommodating the computer equipment, made the human operators very uncomfortable.

Another factor involved lack of sufficient work space for the quality control function. This operation required that the operator compare the original paper document to the electronic image. More space could have provided the operator with a larger, more convenient area to spread out the original file.

Correct lighting settings were important for efficient operations. Lights at a setting that was too bright made it easy to read the paper documents but created a glare on the workstation screens. If the lights were dimmed so that the operators could easily see the screens, it was then too dark to read the documents. Finally, a compromise was reached that afforded sufficient light for comfortable viewing of either the paper document or the workstation screen.

APPENDIX A

OVERVIEW OF DIGITAL IMAGE AND OPTICAL MEDIA TECHNOLOGIES

APPENDIX A. OVERVIEW OF DIGITAL IMAGE AND OPTICAL MEDIA TECHNOLOGIES

A.1 Digital Image Technology

A.1.1 Introduction

Original source documents can deteriorate when used for reference. Therefore, in many cases, it is prudent to convert them to another form or medium that can provide equal or greater utility without harming the original. Usually, a document's value is in the information that it contains. In some cases, however, the document itself holds a significance of its own. Documents containing famous signatures or those that represent an historic event such as a treaty are two examples. Documents that represent a rarity or change in technology such as a particular paper or ink type fall into this category. In general, documents of this type are considered to be intrinsically valuable. Every effort should be made to preserve the original in these cases. However, "information" documents can be converted to another form for ease of reference and storage.

In the past, microform was by far the most popular medium on which to transfer reference copies of original documents. Within the past ten years digital image technology has demonstrated a great potential in many different areas. The basis of digital image technology is that an original document is converted to an electronic facsimile that is stored on a high density digital storage medium and is automatically retrieved to a terminal or printer for reference. Usually the storage medium of choice is digital optical disk. This technology is discussed in later sections.

The combination of these two powerful new technologies offers potential in the areas of storage compaction, image enhancement (in many cases better than the original), automated retrieval, and remote transmission. There are no more out-of-file or lost-file conditions. The reference work is done from the image and not from paper. The following sections will address digital image technology in greater detail.

A.1.2 Document Conversion

In order to convert a document to an electronic image, it must be scanned. When a document is scanned, it is converted into an electronic facsimile. That facsimile or image is then indexed and stored. The index is used to locate the correct image. The image is then automatically retrieved and displayed on a high resolution video monitor or on a hardcopy print.

A.1.2.1 What is a Digital Image?

A digital image is an electronic data file consisting of digital data that when reconstructed either on a monitor or print, appears as the original document. In photography, an analog process, a photo consists of a continuous tonal variety. That is, all information is captured. With a digital image, the image is sampled and converted into the signal much like an image on a television screen. As with a TV picture, the human eye fills in the gaps in information so that it appears like the original.

A digital image is the product of a multi-layered process. The first step is to scan the document. The document will be illuminated with a light from any of several sources. Most scanners use a halogen lamp. Some, however, use low-powered lasers or electron beams in

predetermined patterns or rasters. A photodetector collects the reflected light from the document. Charged coupled devices or CCDs sense the changes in intensity of the reflected light and convert these changes into an analog electrical signal.

In order to understand better this scanning process, a comparison can be made to a window screen. Picture the screen overlaying the document. Each space in the screen represents a single picture element or *pixel*. The horizontal wires forming the screen represent the scan lines. The vertical wires in the screen represent a horizontal row of CCD array elements at the top of the page. The number of scan lines per inch dictates the height of the pixel and the number of CCD elements per inch dictates the width just like the number of horizontal and vertical wires per inch indicate the size of the space in the screen. The more scan lines per inch, the smaller the pixel and the more precise and defined the image. (Note: dots per inch or dpi is used interchangeably with lines per inch.)

Most office documents and many other types of documents and manuscripts require 200 dpi resolution. In some cases, such as in maps, engineering drawings or x-rays, a higher density of 300 or 400 dpi is required in order to capture the minute intricacies in the original. As resolutions increase, the commensurate file sizes also expand to alarming sizes. At 200 dpi, a standard 8.5" x 11" document requires a file of .4675 megabytes. The same document at 400 dpi requires 1.87 megabytes or almost 4 times the size. Large file sizes tax storage systems and data links. Even though at 1,600 dpi the digital image would be virtually indistinguishable from the original, a limit of 400 dpi is more practical since that is the limit of most system peripherals.

A.1.2.2 Scanning From Different Sources

Scanning can create a digitized image from a variety of source materials. As discussed previously, paper documents can be scanned thereby creating a digital image of the original. In a like manner, larger documents such as maps and engineering and architectural drawings can be converted. These larger formats require some specialization in the scanner hardware. In some applications where the presence of these special types of documents is the exception rather than the rule, tiling the document scan is possible. Tiling is the process whereby the operator will scan contiguous pieces of the large original document. The system can be programmed to put the pieces together to form the digitized image representing the whole document. The composite image can then be reduced for retrieval on a high resolution screen or printer. In some cases, large format printers are appropriate for large volume use without the need to reduce the image size.

On the other hand, images stored on microform can be scanned directly from the film in their reduced form. Special purpose scanners are required for this purpose. When scanning from a micro-image, the reduction ratio from the original document size must be taken into consideration. That is, scanning a paper document of 8.5" x 11" size will create a one to one ratio for the digital image. However, if the same document has been microfilmed using a reduction of 48 times (48X), it will require a much higher relative scan density in order to capture the same level of scan density.

A.1.2.3 Scanners

A "scanner" is the hardware component that converts the original to an electronic image. There are three basic types of scanners which are all typically characterized by the movement

of the document and by the location of the optics. Camera-based, flat platen, and moving paper are the main configurations.

Camera-based scanners utilize a series of lenses to focus the reflected light onto a CCD array located where the film would be in a typical camera. The lens/CCD array component is located on an adjustable height tower above the paper. By varying the lenses and distances from the paper, different scan densities can be achieved. This type of scanner is usually used in simple, desktop applications.

Flat platen scanners are contained within a box with a glass platen on top under a cover similar to photocopy technology. The paper to be scanned is placed face down on the platen where it is illuminated by a light source. In some cases, the entire platen with the paper will move over the stationary CCD arrays to facilitate the scan. In other cases, mirrors are used to direct the light and reflection from the paper to the CCD array. Scanners of this type generally have fixed densities since the paper is set at a fixed height from the CCD array.

High speed scanners generally use moving paper systems. In these scanners, the light source and CCD arrays are fixed and sophisticated transport mechanisms move the paper across the scanning field of view. Some scanners of this type can scan both sides of a document on a single pass through the scanner. High rates of speed up to 80 images per minute are possible with some designs.

A.1.2.4 Image Enhancement

While the term "enhancement" is commonly used for the process of "cleaning up" the electronic image, the more accurate word may be to "intensify" the image. Every digital image system is faced with the question of what quality of image will be required for user acceptance of the system. Most systems try to compromise between subjective quality guidelines of just readable to an exact duplicate of every aspect of the original, including stains and damaged areas. One extreme is represented by the full gray scale (discounting color at this time) and the other extreme with black and white. Scan density also plays an important role in the determination. A density of 200 dpi may be just adequate while 400 dpi would give a finer detail. Sometimes the feasible and cost effective approach is to select a lower resolution and level of gray in order to lower the cost per image in a system with a large document universe.

Poor quality original documents can be converted to an enhanced digital image. There are many electronic routines that can be employed. Some are produced by software which are usually slow and not particularly suited for high speed scanner conversion. On the other hand, new enhancement capabilities utilizing computer firmware enable the enhancement process to take place within the time needed to scan a page.

An electronic image can be intensified in a variety of ways. Most, however, have to do with adjusting the contrast between the background and the information on the page. The primary goal of any enhancement tool is to intensify any information or data that exists on the page, no matter how faint, without creating new data. That is, in order to represent the characters on the page accurately, the best possible representation should be made without filling in spots where there is no visible line.

Although there are many sophisticated techniques for this process, thresholding is one of the most beneficial and most straightforward to understand. There are two types of thresholding techniques commonly used. Both constant and dynamic thresholding will be discussed.

When a document is scanned, each pixel[77] on the page carries a number typically between 0 and 255 that indicates its relative level of brightness or shade of gray. [note: some scanners scan as 4 or 6 bits per pixel with correspondingly lower gray scales of 16 and 64]. Since most scanners need to translate this gray scale down to one bit per pixel or black and white [no gray levels], a threshold analysis must take place. With a constant threshold, a value is chosen between 0 and 255 that represents the point at which all pixel levels will be compared. If the writing can be determined to be around gray level of 100, for example, each pixel in the image would be compared to 100. All pixels with a lower number would shift to all black. All above would become all white. (Refer to Figure A-1.)

The constant threshold level is usually variable only as a whole document. This type of enhancement is most useful on documents that have some constant problem present throughout the page. For instance, the entire document could be washed out from exposure to sunlight virtually eliminating any contrast between the writing and the background. In this case, a constant threshold could be chosen that would fall in between the value of the background and the value of the writing. All the writing would turn black and all the background would become white, thereby creating absolute contrast between black and white.

A dynamic thresholding operation is much more sophisticated. Instead of using a single value for the entire page, an analysis is done for each pixel. A threshold value is chosen automatically for each pixel based on its surrounding neighbors. If its neighborhood is generally light, the probability is that the pixel in question should be white and vice-versa. This type of enhancement is especially useful for documents that have inconsistent problems within the same document. For instance, if part of a document is easily readable and another part is very light and difficult to read. A constant threshold setting that would make all the light areas darker would probably totally blacken the darker, easy to read areas. A dynamic threshold would, instead, evaluate each pixel on the page and threshold it individually based on the shading of its surrounding pixels.

In all records management applications, it is important to consider "enhancement" within the confines and requirements of the individual application. When a document is scanned and digitized, it is modified. In other words, all the information that is present on the page is transferred to the image and the extraneous information is discarded. The debate usually centers around what constitutes "information" and who decides what is "extraneous". One of the generally accepted benefits of digital image technology is the ability to clean up a poor quality original document. In the case of a severely stained document, most would argue that to remove the stain is good. There are some, however, who would strenuously argue that the stain <u>is</u> part of the information and should not be eliminated. It is not clear what happens to the argument in the case of a document that is faded to the point that the writing is not readable. Should the image be left in a likewise unreadable condition? Or should the power of digital image technology be used to make that image readable and therefore useful?

---

[77] See section A.1.2.1 on page 165 for an explanation of a pixel.

Histogram

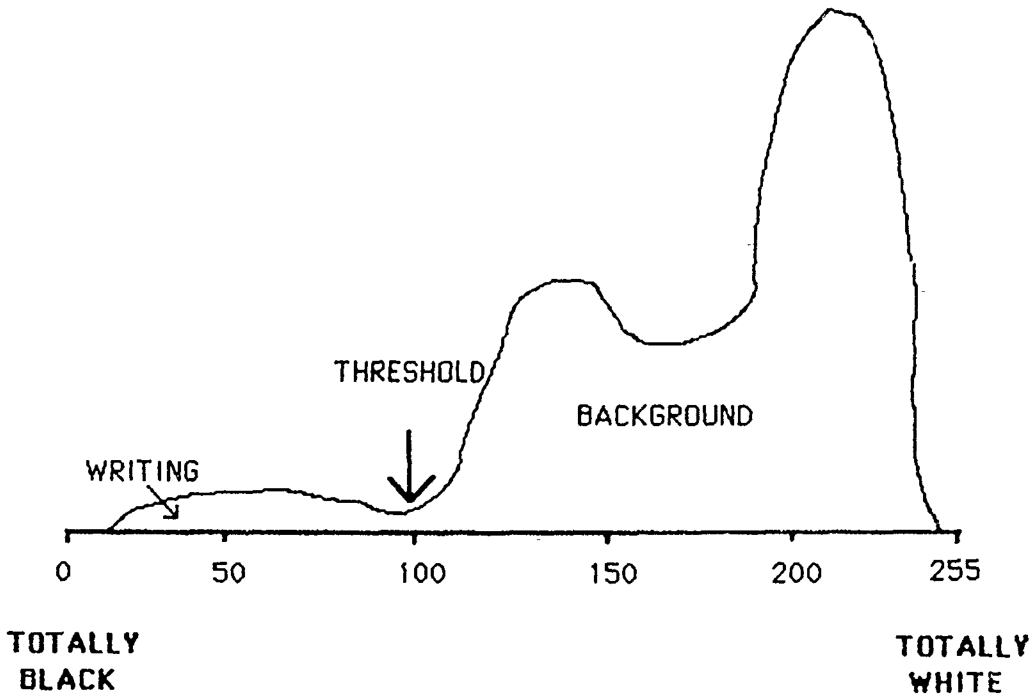


Figure A-1

As noted in Chapter 1, conventional archival wisdom leans toward improved image legibility, and consequently this was the focus in the ODISS project. This report demonstrates that digital imaging technology is a powerful tool for producing highly legible copies of original documents in all but the most recondite situations. Of course, documents of intrinsic value would have to be considered in an operational conversion program. One way to deal with documents of intrinsic value would be to retain the original paper documents and make them available to researchers. An alternative would be to capture images of documents of intrinsic value with full gray scale or color.

A.1.2.5 File Management and Control

In a digital image conversion system, there are two general techniques available to control file management. Both deal with the timing of file identification. The first method does not demarcate the individual logical files until some time after scanning. The second controls the beginning and ending of files during or just before scanning. There are benefits and detriments to both methods.

The first method is primarily used with very small file sizes (less than 4 images per file) in order to speed scanning throughput. If the scanning operation was stopped each time a new file was identified and logged into the system, the production rate would not be optimum. Therefore, documents are scanned in order but without file demarcation in order to maximize speed. Later, operators pull up the images in the order of scan and tag the beginning and ending images of each file. This is sometimes done as part of an indexing process.

The second method is most useful in cases where the file sizes are larger, usually over 15 images per file. (The applications with file sizes that fall between these examples use either method.) In this application, the files are tagged or opened just before scanning their beginning pages. When the next file is opened, the previous file is automatically closed. Using this method, the individual files are defined at the time of scanning.

The size of the file in the application is the determining factor of which of the two methods will be used. In some cases, a combination of both is appropriate. The key to file identification is to make file sizes as large as possible (even artificially) to facilitate the highest possible scanner throughput.

A.1.2.6 Indexing

In section A.1.2.5, techniques for demarcation of files were discussed. Delimiters marked the beginning and ending images of a file. The file, however, had no name with which it could be retrieved. To "index" a file is to attach descriptive information that enables a requestor to identify the file and retrieve it from the storage medium.

Even though large-scale abstractions are necessary for some applications, most only require a simple index of a minimum of data fields. In the case of giant indexes, the data in the index could be greater than the volume of the related image file data. In cases such as these, the application might as well include key entry of all the data on the page and forget about the image conversion, since most of the cost benefit would have already been lost.

The indexing function of an image capture subsystem can take many forms and use any one of several different approaches. These approaches are typically controlled by the factors of operational sequence, data complexity, and input method.

The operational sequence refers to the order in which the index is input into the system. Some applications come with a ready-made, existing database that can be utilized in conjunction with pointers to the related image file. In applications where the index must be created from scratch, the decision must be made as to when the index will be entered into the system. Will the index be created before scanning is done, and then the image files are created and pointers to them immediately placed within the corresponding index entries? Or is the scanning done without an existing index so as to allow for index creation following the image file creation? This second method is especially useful when used in conjunction with the subsequent file demarcation described above. In this case, the operator keys in the index data and indicates the beginning and ending of the file, all in one step.

The data complexity factor is a very important consideration. In most digital image-based systems, the number of index fields is kept to a reasonable number sufficient to provide the searcher with enough information to locate the file without overburdening the database search. The bulk of the research is left to the user to read the image and not rely on the index to supply the required data.

In most applications, the decision is made to have an Ascii-based, free-text search system or a raster image-based system. If a raster-based, "dumb" image system is encumbered with a large index database, search times will be very long and most of the efficiencies of an image system will be lost. Alternatively, image-based systems rely on the researcher to read the document image, thereby reducing the need for elaborate abstracting of the file index.

With key entry operations, the operator keys-in the pre-identified fields either from the digital image or from the original paper document. The lowest point of direct access must carry its own index data. This level is usually the file level in most applications. A "file" can be made up of any number of pages. Individual pages within the file can then be directly accessed after the file is retrieved. In some applications, however, each page is its own file and, as a result, requires that every page be indexed individually.

Optical character recognition (OCR) can be used in some applications for identification and capture of index data. By using this technology, pre-defined document zones are scanned and the raster image data residing there is converted into Ascii character data. Applications that use standard forms or any type of page that contains consistent field locations may qualify for this technology. Currently, standard OCR technology can convert most standard type fonts and some structured hand printing to Ascii code with a high rate of accuracy.

An alternative approach, for those situations that have little consistency in data location, utilizes preprinted labels or header sheets that can be inserted at the head of the file. These labels or header sheets may contain bar code information or character data that are easily converted by OCR. A bar code is a series of parallel lines of varying thickness and spacing that can be read by a bar code reader and converted into character or numeric information. These codes are commonly found on bulk mailings and other items that require automatic high-speed capture of limited quantities of alphanumeric information.

Generally speaking, if it is possible to utilize OCR technology for index entry and is cost beneficial to do so, it should be utilized for speed and ease of use. However, manual key entry can be achieved at extremely rapid rates especially if file sizes are large and entry fields are kept to a minimum.

Another factor to consider for key entry of index data is complexity of data fields. If throughput rate is of concern, as it is in most systems, the index operator should be required to make as few decisions as possible. In other words, the operator should key in the data that is called for and not have to compose an abstract of the information in order to fill the field.

A.1.2.7 Quality Control and Assurance

Most designers of digital image-based systems recommend that some type of quality assurance operation be built into the system design. This is particularly true since most of the systems of this type employ write once digital optical disks as the "permanent" storage media of choice. Since information is not easily modified on these disks, all corrections should be completed prior to writing to optical disk.

Quality control and assurance can be split into two main categories. The first is quality control of index data. The second is image quality assurance. Each has a place in every system. The only question is one of the extent of coverage. Are 100% of all files verified for accuracy in all data fields? Or, is it more prudent for the particular application to check a sample of perhaps only 10%?

It is important to perform a dual check of images. Each image should be matched against the original to verify that it had been scanned. As this check is being conducted, image quality is screened as well. (Refer to section A.1.2.4 on page 166 for a more complete discussion of image quality and analysis.) Images not meeting quality standards are rejected and usually sent to be rescanned. Conversion subsystems using a table top scanner usually combine scanning quality control and rescanning intc one operation for convenience. Conversion subsystems using a high speed scanner usually separate these activities in order to streamline the process. Consistent speed and routine are upset when a document is rescanned in a high speed operation. When high speed scanners are the primary conversion tool, rescanning is usually done with table top scanners that can have more flexibility with their scanning functionality. Image quality control and image rejection in a high speed system are generally a "tagging" process. An electronic tag is associated with a bad image to facilitate locating it later for rescan. When the file is ready for rescanning bad images, the tagged images can be brought to the screen and the original documents rescanned with more finesse than is possible with a high speed operation. The newly scanned, better image is then substituted for the poor one in the image data file.

A.1.2.8 Data Compression and File Size

Digital image systems produce stringent demands on data transfer, processing, and storage systems because of the very large file sizes common to image data. Since the image file sizes are so large, various methods may be employed to reduce or compress the size of the file without noticeable data loss. Bitonal or black and white images contain long strings of 1's and 0's indicating areas of black and white on the image. Typical office documents contain quite a lot of white background space. Therefore, it is possible to combine long strings of identical data in order to create smaller data files. For example, if the next 12,000 pixels in rows were white [or 0], that data could be described as a byte of data representing 12,000 white pixels.

That example would demonstrate a savings of 8 bits of data for the compressed string compared to 12,000 bits of data for the uncompressed string. This method is referred to as

run-length encoding, and is commonly used in many applications. This method will typically convert an 8.5" X 11" business document scanned at 200 dpi with one bit per pixel (black and white) from its uncompressed form of 4.68 megabytes to a compressed form of around 50 kilobytes.

Other compression techniques have been developed to handle cases where there are no very long runs of duplicate pixels. Huffman encoding utilizes an algorithm that predefines common pixel sequences and stores them away. When the pixel run matches one of these sequences, a code is inserted in its place, thereby, saving space relative to the size of the substituted string. There are many custom compression algorithms that can yield extremely high compression ratios.

Obviously, some types of images compress at a higher rate than others. A document with a high percentage of white space will produce a very high compression rate and a small file size. Alternatively, a photograph may not effectively compress at all since there may be little redundancy in the image bit string.

The Consultative Committee on International Telephones and Telegraphy (CCITT), an agency of the United Nations, has developed international standards for the transmission of facsimile digitized images or FAX. These methods, known as Group III and Group IV FAX, use run-length and Huffman encoding as the basis for their process. Both groups are defined for digital images. The main difference is that Group IV FAX uses a two dimensional analysis instead of Group III's single dimension. Table A-1 shows relative file sizes with typical compression rates.

Image Compression and File Sizes for an 8.5" X 11" Office Document

DOTS PER INCH	UNCOMPRESSED	10 to 1	15 to 1
200	470 KB	46.8 KB	31.17 KB
300	1051 KB	105.1 KB	70.13 KB
400	14,960 KB	1496.0 KB	997.30 KB

Table A-1

An image is usually compressed soon after scanning in order to limit its burden on the rest of the system. In most systems, the compressed image is stored on the capture storage buffer and is transported around the system as a compressed image. The only time that the image is decompressed is when it is prepared for display or printing.

A.1.2.9 Image [Input] Data Buffer Storage

In all but the smallest, most simplistic systems, magnetic disk buffer storage is used as a temporary location in which to store scanned images awaiting preparatory operations for writing to a non-rewritable, long-term storage medium, such as write-once, digital optical disk. As the images are created by the scanner, the image files are compressed and are written to magnetic disks. Throughout the processes of indexing, quality control and rescanning, the images will reside on this magnetic buffer so that they can be modified if

necessary, as in subsequent rescan and image enhancement. Once all actions are taken on the image file, it is written to the long-term storage medium. In some cases, magnetic tape is used for the temporary storage of digital images. This technique is common in very large, high speed operations that require a longer temporary storage period until the files are transferred to their long-term storage location. This situation would cause a large build-up of temporary data and would make magnetic tape an attractive alternative to magnetic disk because of a lower cost per bit.

A.1.3 Image Retrieval

Image retrieval is the user end of the system. Usually, it is the only point of contact that the researcher has with the entire system. It is for this reason that the user utility of the retrieval subsystem can be the sole basis of acceptance or rejection of the complete digital image system. If the user finds fault with any aspect, he may attempt to return to the manual, paper-based system with which he is most familiar.

Image retrieval includes identification of the image file and creation of either a hard or soft copy (i.e., screen display) for viewing. The following sections will describe the process of image identification and image output onto high resolution screens and laser printers.

A.1.3.1 Locating the Image File

In order to locate an image file, an index search must be conducted. This index must have already been created and implemented. Database software is commonly used to manage image file indexes. These indexes typically use key data fields that can be searched by using various boolian search procedures allowing for "and" and "or" conjunctive conditions between fields. An example might be to search for all last names of Smith and all first names of Joe or John. Many indexes allow for word truncation and wild cards. In the example above, if we did not know exactly how to spell Smith, but we were sure it started with "Smi", we would key-in Smi* to select all words in that field that begin with "Smi".

The results of a successful search [when at least one file meets the search criteria] are displayed in a list form, in most cases. The researcher would then choose from this "hit" list to pick the file he would like to print or see on the screen.

A.1.3.2 Image [Output] Data Buffer Storage

When digital optical disks are used for long-term image storage, a temporary magnetic disk buffer is typically used to assist in retrieval. When the request comes in, the images are read from the optical disk and spooled off to the buffer as an interim stage. Once the file or set of requested images is present in the buffer, the optical disk is again free for subsequent retrievals. The image file is sent to the workstation or to a print server. In order to facilitate a faster response time to the first image, some systems will display the first image in the file while the others are being transferred to the workstation. This will enable the user to begin reference at the earliest time period and with the shortest wait time.

At least two types of image buffers are used on the output subsystem portion of a system. The primary buffer is the one previously mentioned. The other is the cache buffer on the receiving workstation and works both to facilitate a faster "page turning" response at the workstation, and to have a place to store the image while the decompression is taking place. Either magnetic hard disk or ram storage can be used for this purpose.

A.1.3.3 Image Workstations

The image workstation is defined as any terminal which has a screen that can render a representation of a document image. An image workstation is the primary reference tool in a digital image-based system. Because a digital image can be effectively displayed in a manner that makes it appear as a facsimile of the original paper document, a paperless reference system is theoretically possible. In reality, however, many users will still request hardcopy prints (if available), at least until they are comfortable with the transition to an entirely new way of working that no longer requires the traditional reference methods using paper.

There are five major categories of screens or monitors that can be used in conjunction with a digital image system. They are: electroluminescent displays, light emitting diode displays (LED), gas plasma displays, liquid crystal displays (LCD) and cathode ray tubes (CRT). Most systems utilize cathode ray tubes for their image workstation monitor screen.

A.1.3.3.1 Display Density

One of the primary concerns of system designers and users alike is the determination of the display density of the screen. Display density is similar to scan density since each describes the number of scan lines that define the image. The higher the number of lines (or dots) per inch, the more well-defined the image. As the display density increases, so does the cost of the terminal.

Regular monochrome monitors may have a cost of less than $100. They do not, however, have sufficient display density or resolution to display an 8.5" x 11" office document image in a way that will show a completely legible image at full size. Nevertheless, monochrome monitors are used in low-end systems as image terminals. In these applications, images are displayed at less than full size as a whole document. At this resolution, the document is not readable. Zooming of a particular portion of the document image is necessary in order to make it readable.

Typically, monitor screens chosen for image based systems are large enough and have sufficient resolution to display an 8.5" x 11" office document in a manner that is readable when displayed at full size. This usually requires a 100 lines-per-inch display density both horizontally and vertically. In many cases, this too requires frequent zooming of parts of the image to render it readable. The most common density is 150 lines per inch, both horizontally and vertically. At this density, virtually all documents containing characters above 4 point size are easily readable, even without the necessity of zoom.

Unfortunately, as a monitor's screen density, size, and other capabilities and features increase so does its price. In applications that require a large number of image retrieval workstations, the costs may prohibit use of higher density monitor screens since the cost differential may be as high as 700% between 100 and 150 lines-per-inch capability.

A.1.3.3.2 Simultaneous Display of Image and Character Data

Another important feature of image monitor screens is the capability to display simultaneously image as well as character data. In the past, most systems required two monitors on the desk. One was a regular monochrome or color monitor used for display of character data, such as index and menu information. The other was the image monitor

screen. Even though many applications still use this configuration, other applications are choosing to use larger sized monitors that have the capability to display both character and image data at the same time, on the same screen. Costs usually dictate the decision on which way to proceed. Applications that already have a large installed base of personal computers may choose to integrate an "image only" monitor screen. With this configuration, they could use their existing system for index and menu operations and the new screen for image retrieval (Refer to Figure A-2). This is usually a very cost effective approach.

Applications that either have no existing workstation or are small enough to justify a higher cost monitor screen may choose to use a larger screen with the capability to display both image and character data simultaneously. (Refer to Figure A-3.) With a full-sized page and character data to display, screen size can be an important factor. Screens that are set up in portrait mode (with longest side at the vertical) that are exactly 8.5" x 11" in size usually require character data to be displayed in windows overlaying the image. If the image is not to be obscured by character data, the screen must be larger. Most screens of this type are 19" or 20" measured diagonally.

A.1.3.3.3 Display Featur

There are a number of things that can be done to the image in order to increase its utility to the user. In most cases, the image is temporarily modified for the display only and does not result in permanent change to the image. Usually, permanent changes take the form of image editing. Image editing is the term used to define the addition or removal of pixels from an image. Temporary modifications, which typify display features, will be discussed here.

<u>Image Sizing</u>

Image sizing normally refers to the increase or decrease in the size of the image relative to the screen. Images too large to be displayed on the screen can be scaled down in order to fit the screen dimensions. Normally, large images are reduced in size equally in both the horizontal and vertical dimensions so as not to distort the images. This is usually accomplished by pixel reduction. With this technique, pixels are eliminated in a predetermined sequence that would result in the required scale of the image. If the image needed to be reduced by 20%, every fifth pixel along each display row would be eliminated as would every fifth row.

Another method for display of images too large for the screen without reduction uses scrolling and panning the image. In this case, the image retains its original scale. The screen can only display a portion of the image at a time. The operator can then pan horizontally across or scroll vertically up and down the image using either a mouse or cursor keys.

The opposite sizing technique is image enlargement or zooming. This is a very popular and useful asset to a terminal's capabilities. Most image terminal screens do not have the resolution capability to display at the same density of the original scan. If an image were created at scan at a resolution of 200, 300 or 400 dots per inch, it would probably be displayed on a terminal screen capable of 100 or 150 dots per inch. In other words, all of the data from the scanned image would not be displayed. In order to display all the scanned image data, a portion of the image can be displayed at the original density. The size of the portion depends on the resolution of the screen and the density of the scan. Each pixel that was created will be displayed. If an image were created at 200 dots per inch and displayed

Image and Character Terminals

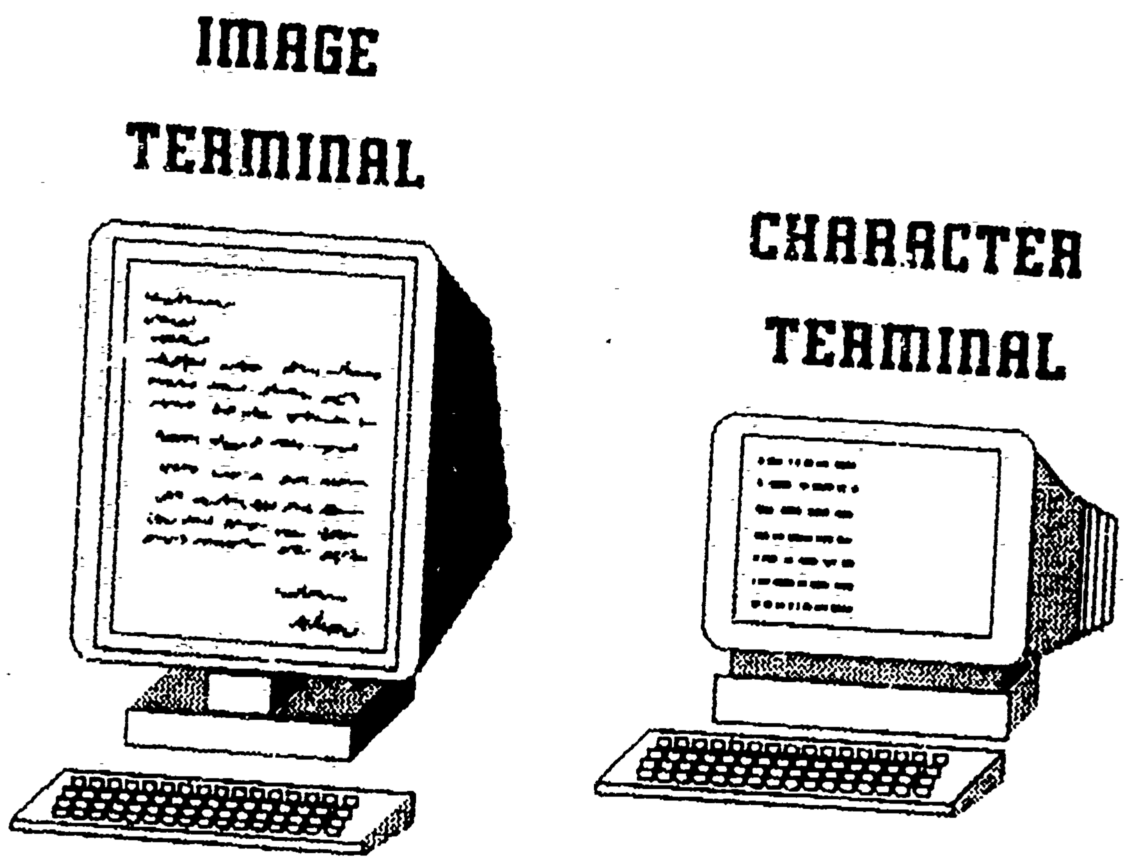


Figure A-2

Image Terminal

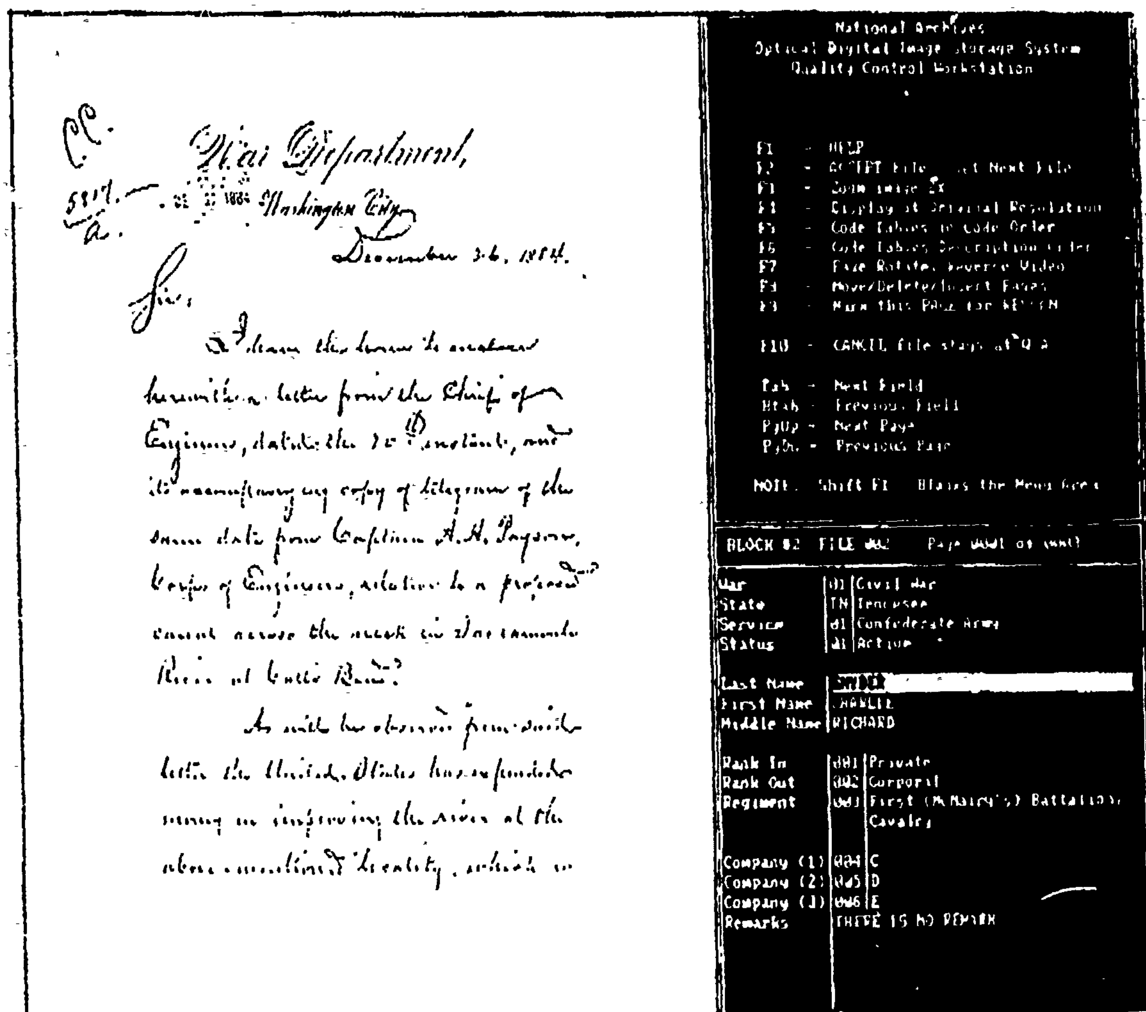


Figure A-3

on a 100 dots-per-inch screen, the image could be enlarged by a factor of two in both horizontal and vertical dimensions in order to effect a zoom factor of 4 and display all original pixels.

It is also possible to expand the size of the image on the screen by various factors. This is done by adding redundant pixels and does not increase the amount of real data gleaned from the image. It can be useful, however, for simply increasing the size of the image for viewing.

Other Features

Several other image monitor features are notable for their user utility. Image inversion, image rotation and screen printing are all common features.

Image inversion is useful when the operator wishes to change from black characters on white background to the opposite white on black. This is especially useful when viewing images scanned from a negative microform. The process is fairly simple since the image source is digital ones and zeros. If a positive image carries a one equal to black pixels and a zero equal to white pixels, to invert the image only requires the pixels to be reversed.

Many applications use images that may be oriented in both landscape (horizontal) and portrait (vertical) modes. Image rotation allows the user to pivot the image either a predetermined or a variable number of degrees. This capability can be accomplished either by a screen that physically rotates or, more commonly, by electronic means. This requirement is usually stated as the ability to rotate 90 degrees left and right and 180 degrees.

Screen prints are used when the operator wants a "snapshot" print of the screen's contents. This capability can usually be exercised at any point in the process. Users find this especially useful when using a large screen with simultaneous index and image capability. One print can show the document and informational data on the same page.

A.1.3.4 Printers

Virtually all digital image systems require the capability to output on hardcopy. Laser printers are by far the most common type of printer used for this purpose. Laser printer technology is very similar to that used in electrostatic copiers and represents a well established, mature component of the system.

Laser printers, like photocopiers, use either a laser or other light source to create a transient image on a photosensitive surface. This transient image is developed by applying toner. It is then transferred and fused to paper with high heat. This process is repeated for every new image printed.

A.1.3.4.1 Print Density

Just as in the case of scanners and terminal screens, density is very important in image printers. Virtually all printers are based on a resolution of at least 300 dots per inch. Some new designs have capabilities of 400 dots per inch and higher. If the scan density matches the print output, the print will simply be on a one-to-one scale. However, if the scan resolution is different, the image must be scaled to accommodate the print resolution.

A.1.3.4.2 Print Speed

Print speed is stated in pages per minute. Low-end printers are rated at seven to eight pages per minute. Mid-range printers reach twenty pages per minute. And high-end printers can print over one hundred pages per minute. Printer speed is a function of many factors, not the least of which is image buffer storage. The printer can print only what it has on hand. The greater the storage buffer space, the greater the number of images ready to print and the faster the print speed.

A.2 Optical Media Technology

A.2.1 Introduction

Digital image systems require very large storage capacities due to the considerable sizes of image files. Along with other capabilities, digital storage media must have a low cost per bit to qualify as a viable medium for these systems. Digital optical disk technology offers a good solution for the long-term storage requirements for digital image-based systems. This section will describe the wide variety of media types, recording methodologies, and other factors that should be considered in the selection process.

A.2.2 What is an Optical Disk?

The historical beginnings of optical disk technology had their roots back in the 1930's with some initial experiments. An assortment of techniques has been attempted over the years with varying degrees of success. Most recently, the technology was split in the early seventies with what was categorized as the contact and non-contact techniques. Capacitance Electronic Disc (CED) and Video High Density Disc (VHD) systems both used a contact method with a stylus on grooved and grooveless disks, respectively. These techniques were akin to audio records. By far the most popular, however, were the non-impact methods established at about the same time. These used a laser to create and interpret the information on the surface of a disk. Since the non-impact method has survived and flourished, it will be the subject of this chapter.

An "optical disk" is a disk that stores analog and/or digital data and is optically "read" by a laser as the disk is spinning at high speed. The term "videodisc" generally refers to optical disks that store *analog* data. "Optical disk" is used for *digital* data disks. The main distinction of both types of disks is their ability to store a vast amount of data in a very compact space. Analog videodisc technology is mainly used for still frame photographs and motion video storage. Digital optical disk technology is the primary long-term storage medium for digital image-based systems.

The refinement of laser technology has had an important effect on the development of optical media. The laser beam is used to write to and read from optical disks. In the past, gas generated lasers were used in these systems. They were very expensive, difficult to maintain, and had a very short life. They were also large and unwieldy. The development of the semiconductor diode laser changed the entire industry. It now had a laser that was the size of a pencil eraser, inexpensive, and was long lived. Based on this and other related developments, compact disk audio or "CD" has taken the marketplace by storm. It is hard to believe that an optical disk, spinning at high speed and being read by a sensitive laser beam generating music can be taken along while jogging.

Some types of optical media are strictly for playback and must be mass-produced in a factory. Others can be written within one's own system. Both types of recording methodologies and other characteristics are discussed in the next section. All types of optical recording media use a laser to read changes in the light intensity of its reflection. This data is converted to either an analog or digital signal which carries the recorded information.

Each type of optical medium shares at least one component with its cousins, an optical block. The optical block (see Figure A-4 and Figure A-5) is the unit that carries the laser source, lenses, and sensors that read from and write to optical media. The optical block also acts to focus and track the laser along the spinning disk. It is one of the key components in any disk based optical memory system.

A laser source, usually semiconductor diode laser, is used both for writing [write once and rewritable media] and for reading information. The laser beam is focused, through prisms and lenses, and shown on the highly reflective surface of the disk. In order to write information on the disk, the beam is modulated to a higher power and through this higher power changes the reflectivity of microscopic portions of the disk. To read the information, the laser power is reduced and the beam is shown on the disk again. This time, the minute changes in the amount of reflection being sent to the sensors is translated into data. There are a variety of techniques for recording information on a write once or rewritable optical disk. These will be discussed in the next section.

Grating is the term used for splitting the laser beam into three separate beams (see Figure A-6). Generally, this function is performed in order to use the two outboard beams to regulate and steer the central or information-carrying beam along the track. After the beam has been split, it goes through a polarization beam splitter (PBS). A PBS contains a dielectric multilayer that permits the beam to pass through or deflects it to the receiving sensor depending on the direction of polarization of the beam (see Figure A-7). The beam, at the laser source, is horizontally polarized. The PBS allows it to pass straight through to the disk. When the laser light is reflected from the surface of the disk, it travels through a quarter wavelength plate (QWP) that turns the horizontally polarized beam to a vertical orientation (see Figure A-8). That vertically polarized beam travels back to the PBS where it is deflected ninety degrees since it is now vertical. The photo detector sensor receives the beam and converts it to information. If this polarization technique were not used, two separate lasers would have to be used.

Another very important component is servo circuit. Servos control movement of the optical block and the rotation of the disk. The tracking servo uses the two outboard beams that were split off from the laser source by the grate to control the movement of the optical block and to ensure that it stays along the disk track that carries the data. As the "F" beam and "E" beams move off the track, they will send messages to the optical block to move to trace the track accurately (see Figure A-9).

The focus servo keeps the distance constant between the object lens and the disk surface. These changes can be caused by disk irregularities and disk flutter as the disk rotates at high speed (see Figure A-10). The photo detector sensor can recognize subtle changes in the beam shape caused by changes in the distance of the laser source from the disk surface (see Figure A-11). When these irregularities are found, the optical block is adjusted to compensate. There are other servos, as well, that handle such things as rotational speed of the disk and regulation of the skew of the optical block.

Optical Block

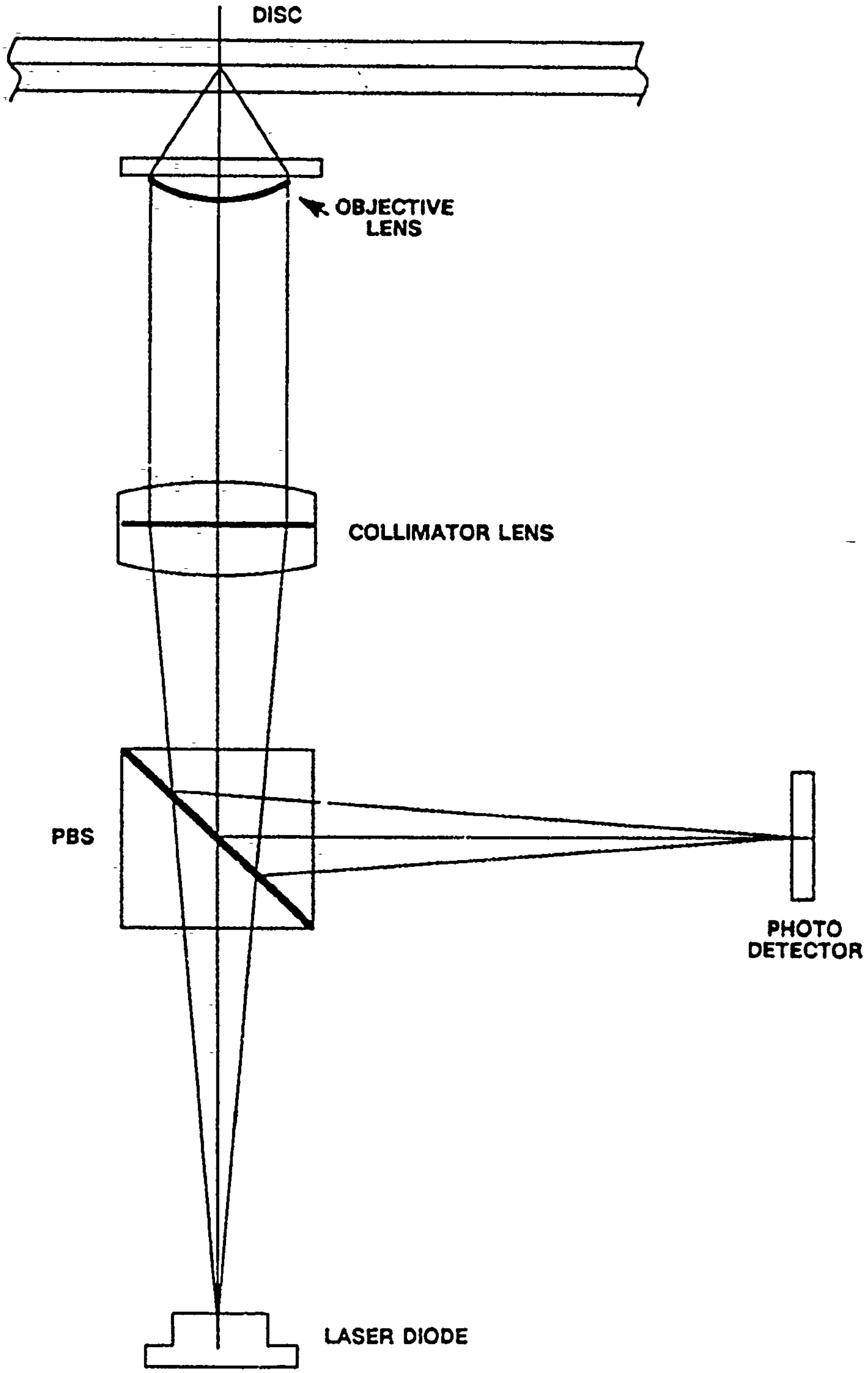


Figure A-4

Optical Block Mechanism

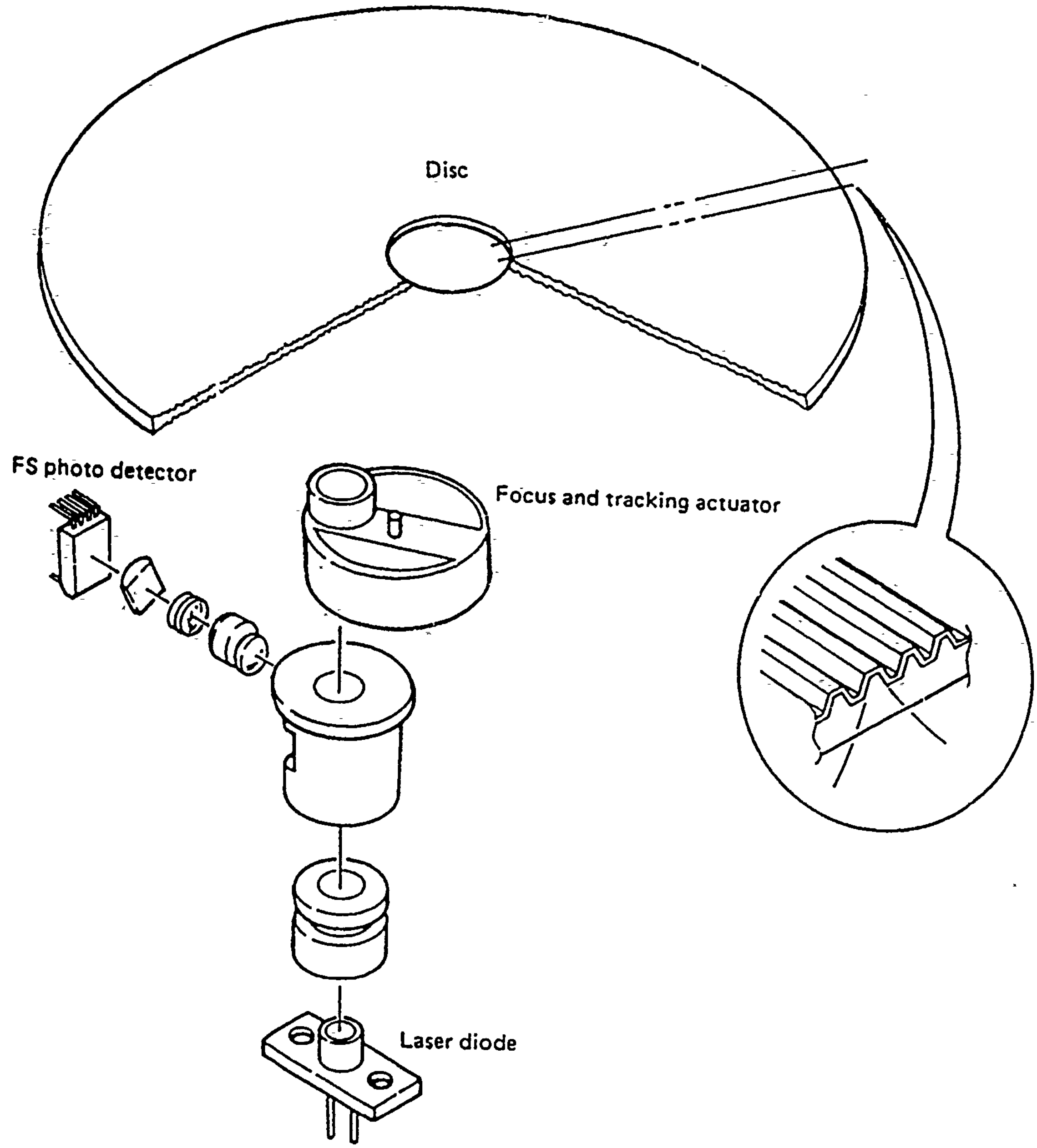


Figure A-5

Beam Grate

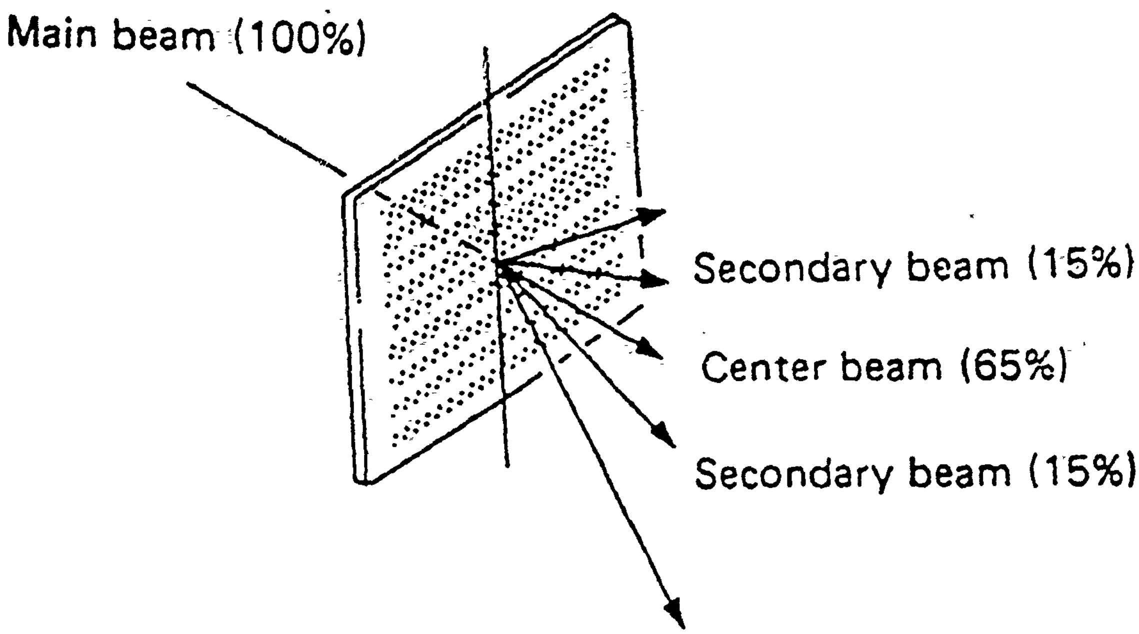


Figure A-6

Principle of PBS

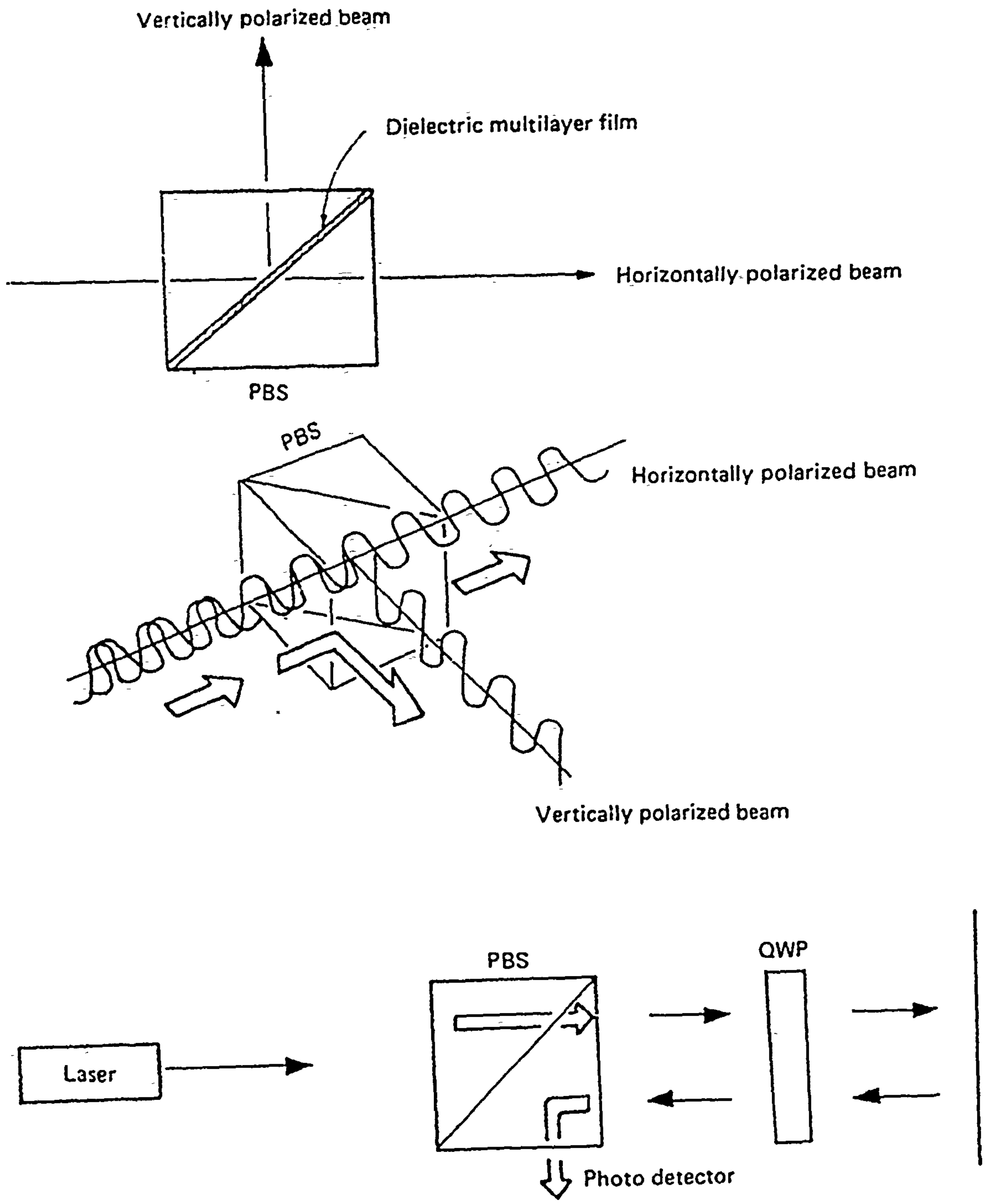


Figure A-7

Role of PBS

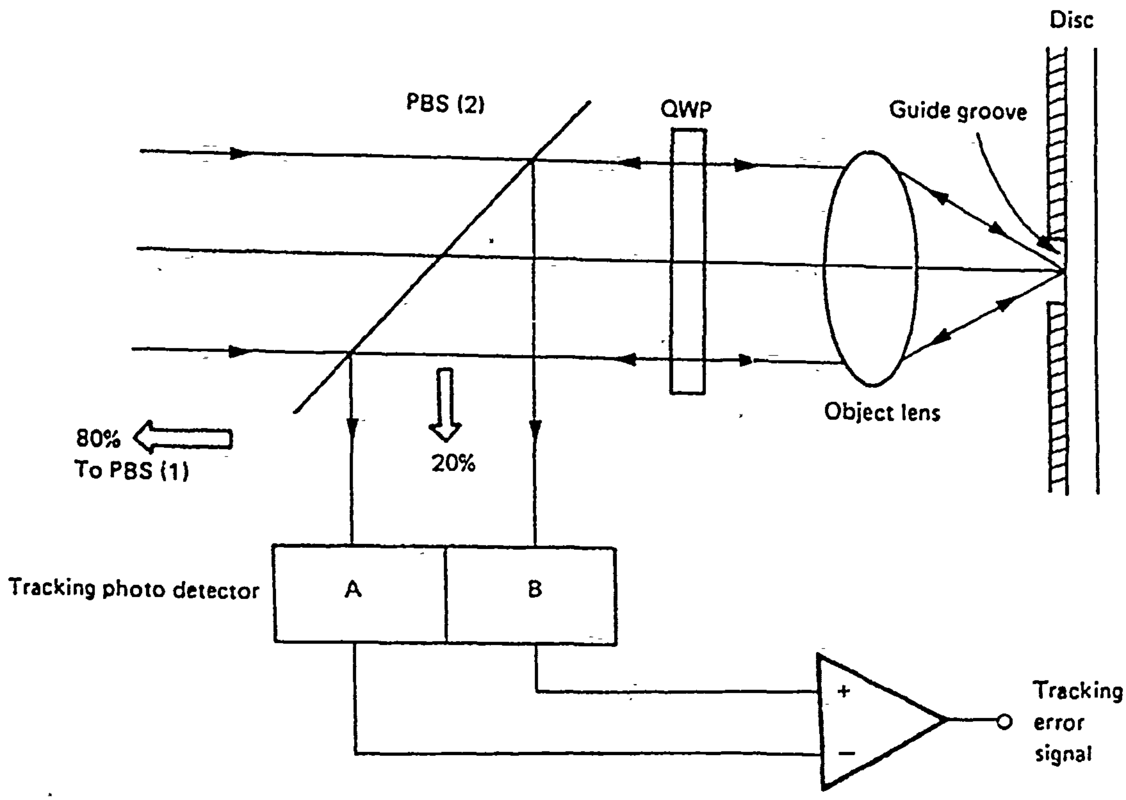


Figure A-8

Principle of Tracking

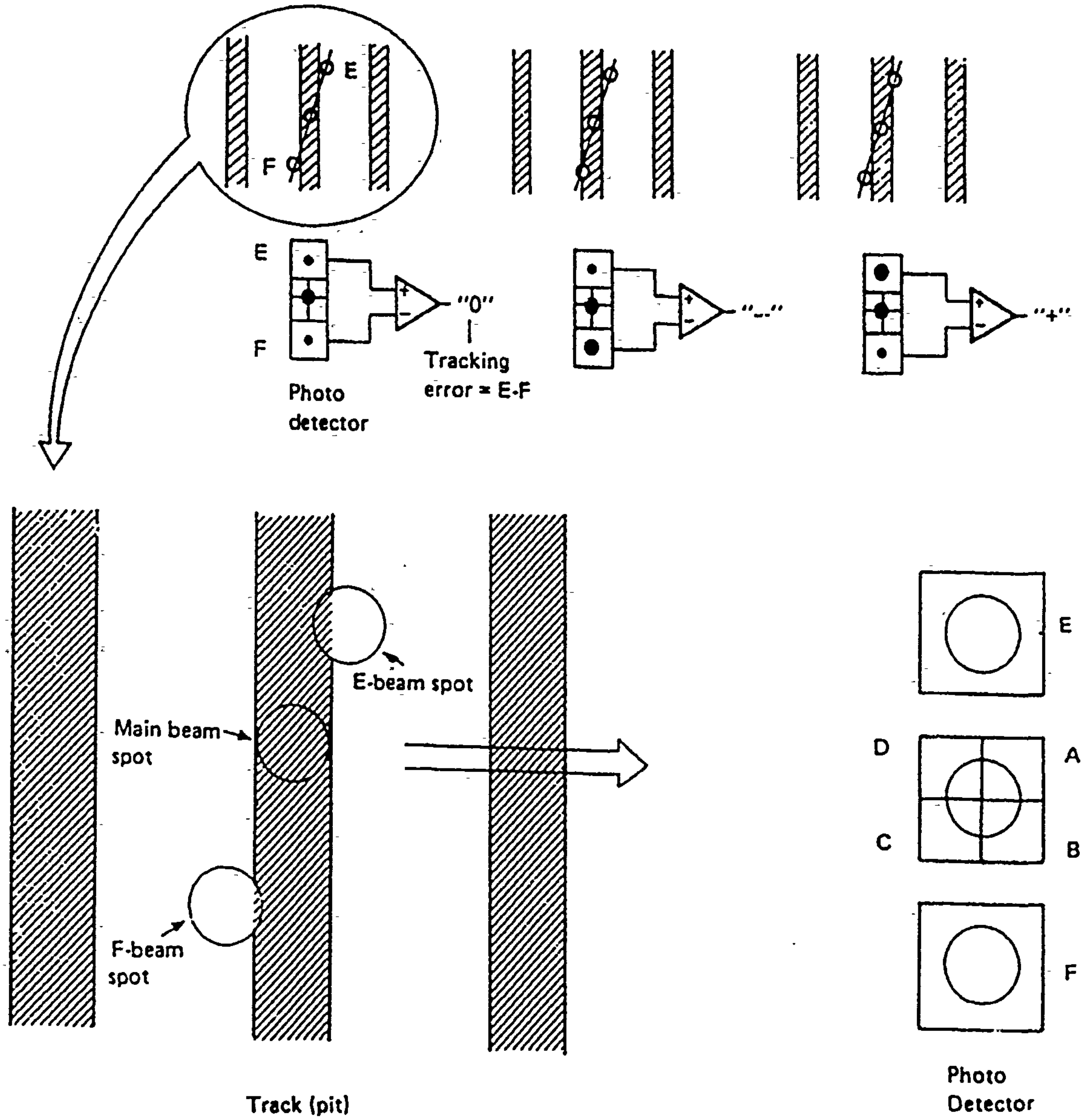
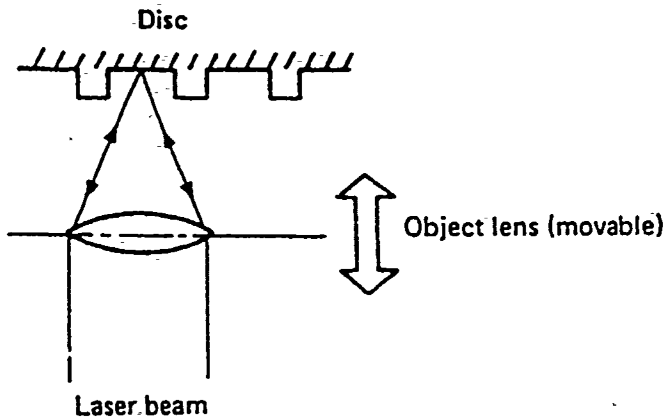


Figure A-9

Principle of Focus Servo



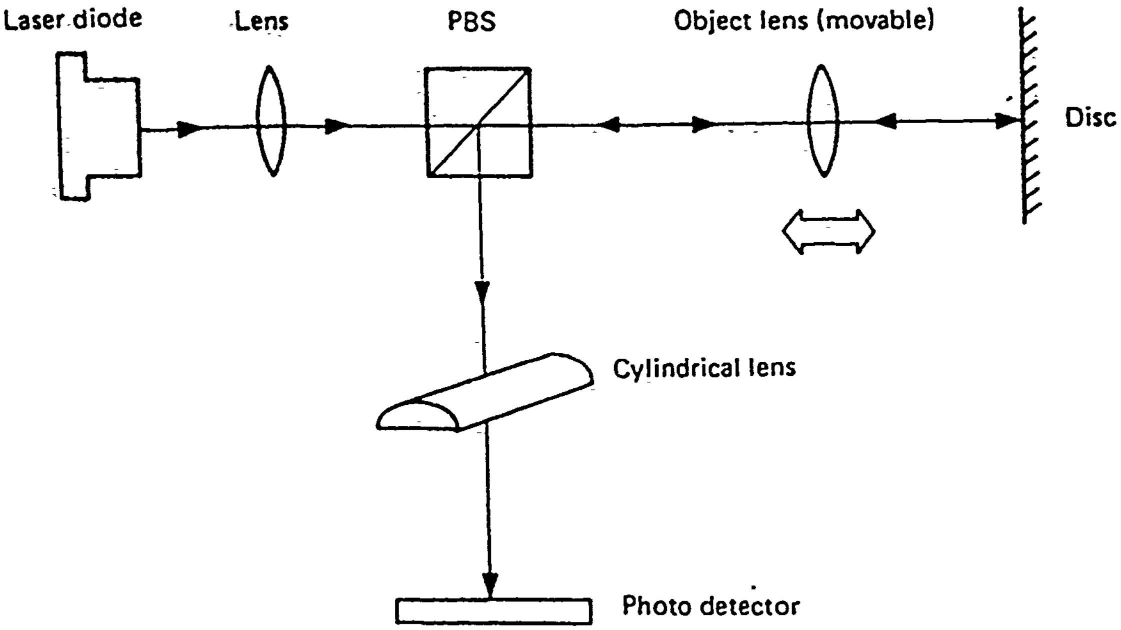


Figure A-10

Principle of Focusing

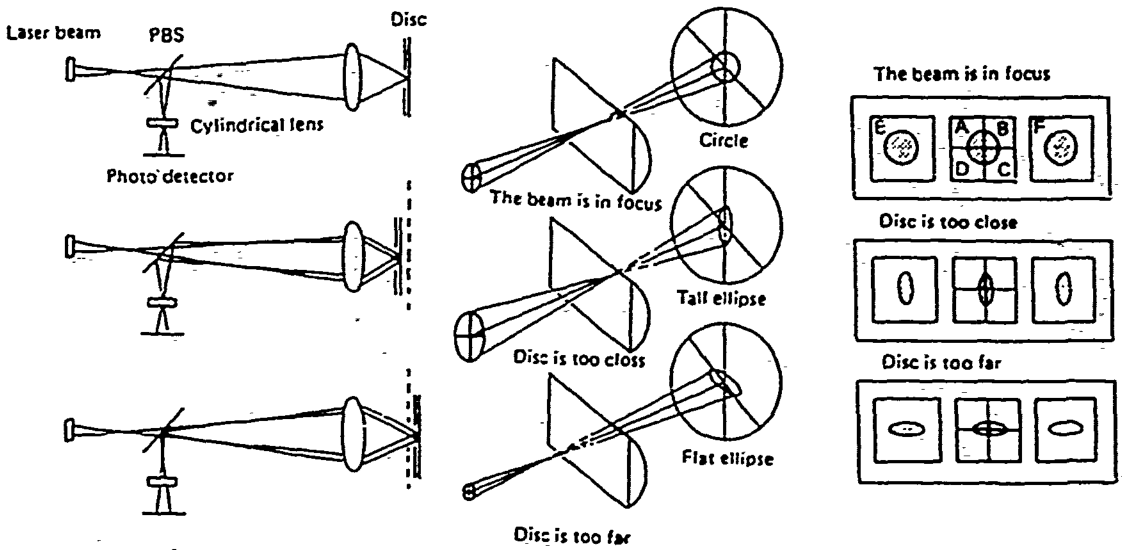


Figure A-11 Principle of Focusing

The principles of recording and playback of optical disks are basically the same. The disk is made up of several layers of material (see Figure A-12 and Figure A-13). These layers usually consist of a highly reflective metallic layer that carries the data, a surface substrate and a protective layer between disk halves. A laser beam is directed onto the reflective layer of the disk. Changes in the intensity of the reflection of the laser beam are interpreted by a sensor and converted into electrical impulses. These impulses carry the information. There are differences in the various types of disks in terms of their recording techniques and reading methods. These differences will be discussed as these formats are reviewed in the following section.

A.2.3 Optical Storage Formats

A.2.3.1 Analog Videodiscs

There are a number of distinguishing characteristics of analog videodisc technology. They can be categorized by recording methodology, primary applications, and limitations.

Videodiscs, as defined here, store analog information. Analog information is carried on a signal that continually varies according to the range of intensity and frequency. Digital information, on the other hand, can be defined as a discrete, off-and-on signal. This concept may be illustrated by thinking of a light dimmer used to regulate gradually the brightness of the light as an analog process. Digital could be presented as an off/on switch. The electricity is either off or on with no varying levels in between.

Videodiscs are in a category of optical disks known as ROM disks or Read Only Memory. The recording and creation processes of virtually all types of ROM disks are similar. The data chosen to be transferred to a ROM disk must go through a premastering process where the information is put into the correct format and prepared for mastering onto the disk. Next, the actual mastering takes place where the data is converted to a one-inch professional videotape and sent to a factory for disk creation. At the factory, a stamper disk is made that will be used for creation of the many replicate disks in much the same manner as an LP record album is produced (see Figure A-14).

Videodiscs can effectively store and reproduce analog audio or video signals at a high-density rate. The typical videodisc can store 54,000 separate frames of video, each representing a single photographic print or slide. Since the videodiscs are almost always two-sided, the total frame capacity equals 108,000. The video tracks are supplemented with two audio tracks that can be played simultaneously with the video. Each frame can be viewed constantly for an unlimited time with no degradation to the disk. This stop action feature is not possible with videotape for any extended length of time. Full motion video is also possible with videodiscs. In this case, the frames are shown at the standard rate of thirty frames per second. Videodiscs have the capacity to hold enough information for from one to two hours of full motion video.

Analog videodiscs are best suited for audio, video, and photographic reproduction since the analog signal produced from a videodisc is compatible with standard television signals. These applications usually do not require high levels of resolution or definition. Typical applications include: training, motion pictures, still photography, and educational activities.

Write-Once Disk

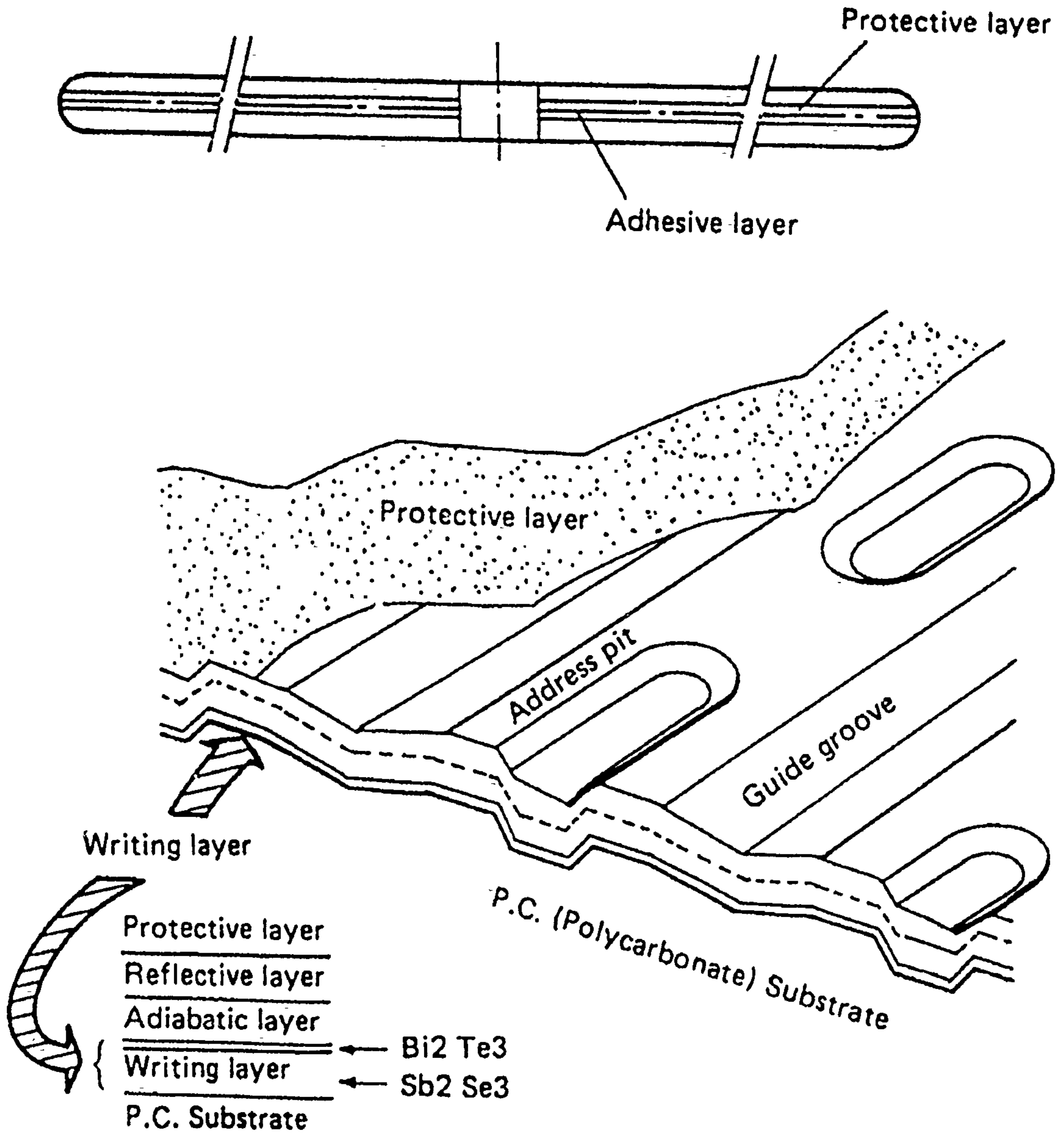


Figure A-12

Videodisc

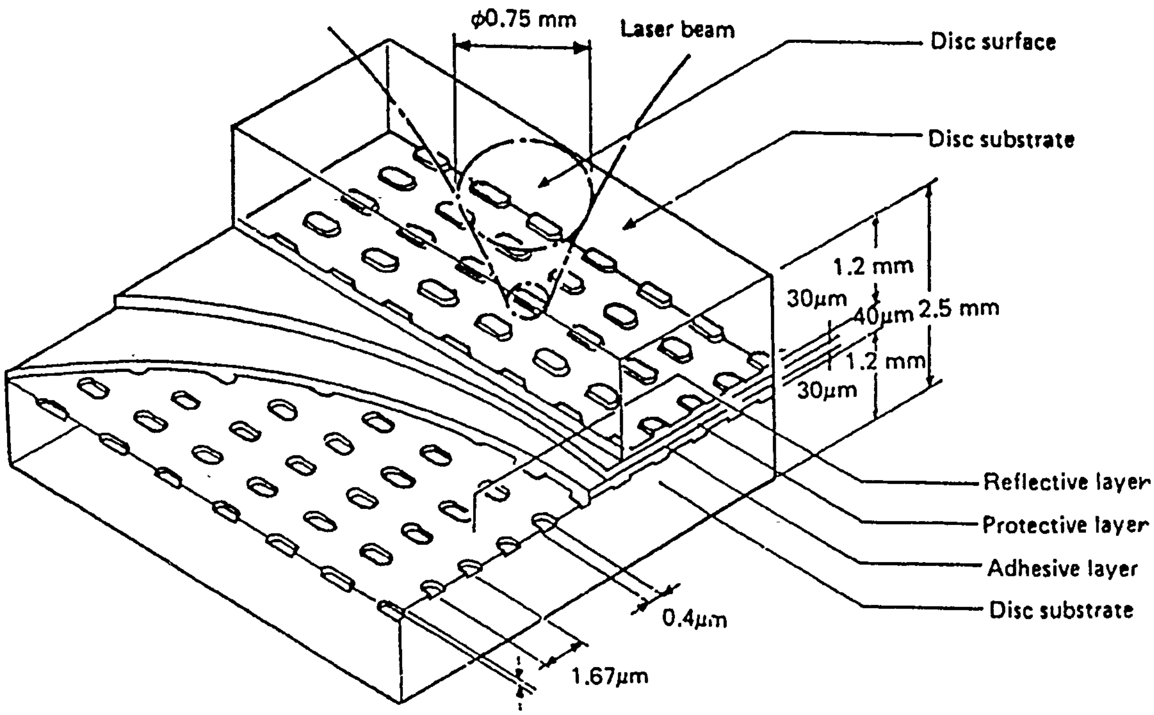


Figure A-13

Videodisc Production Sequence

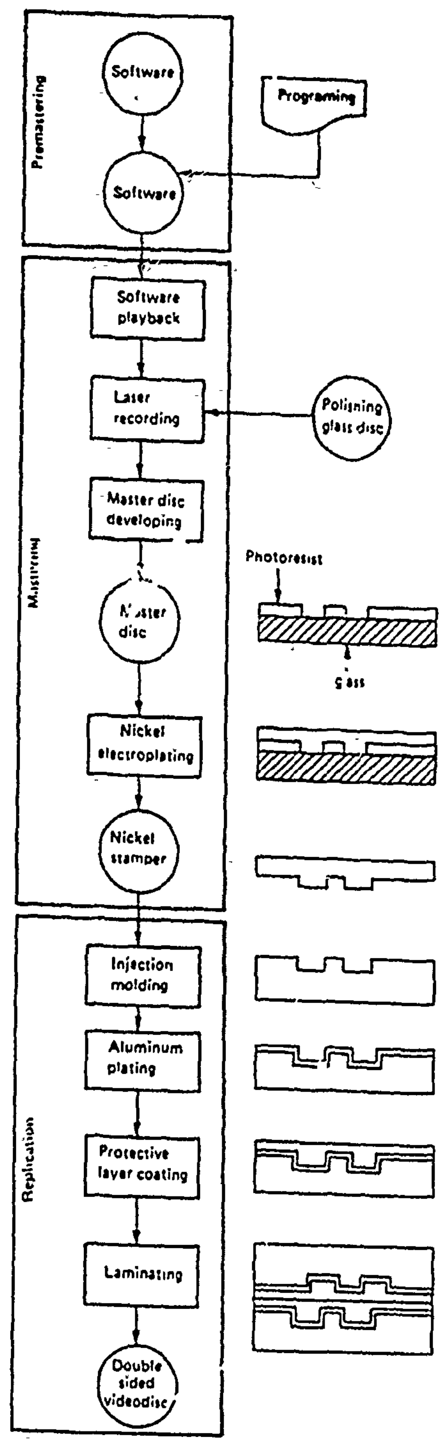


Figure A-14

A.2.3.2 Write-Once Digital Optical Disks

Write-once, read-many or WORM optical disks were the first type of digital optical disk to become commercially viable. The term WORM signifies that the disk can be written similar to a magnetic disk by using a disk drive without the necessity of being mass produced in a factory. However, unlike magnetic disks, WORM media are not rewritable. That is, a new set of data cannot overwrite existing data on the disk. In most archival settings, this characteristic is useful since the recorded documents rarely require updating.

There are a variety of sizes of WORM disks ranging from 5.25 inches to 14 inches. The two most common sizes of WORM disks are 5.25 inch and 12 inch. Both are used in a variety of applications from personal computers to massive mainframe based systems. The smaller 5.25 inch disks are the fastest growing segment of the WORM market. Their storage capacity ranges from 230 megabytes to 1.2 gigabytes. Since the disk drives are the same size as a floppy disk drive, and since the media are removable, 5.25 inch disks are very popular in personal computer based systems. WORM drives in this size usually range in price from $1800 to $5000. The price per disk, in quantity, is from around $100 to around $360.

There are a number of reasons for the great increase in 5.25 inch based systems in the last couple of years. The following are just a few of the most important considerations. 5.25 inch disks are much cheaper per disk than 12 or 14-inch WORM disks. Their drives are cheaper as well. Their capacities have grown to the level of 12 inch disks of only a few years past. Their mass is much less than the larger formats allowing for much simpler and less expensive jukebox storage units. And finally, they are much closer to having governing national and international standards.

The 12-inch WORM marketplace consists mainly of moderate to large-sized systems requiring very large amounts of data storage. There are at least six major manufacturers of 12 inch WORM disks. There is little compatibility or standardization in this medium. Capacities range from 2 to 6.8 gigabytes. The next generation of media will offer double density with a resulting lower price per bit.

Kodak is the only manufacturer currently offering a 14-inch WORM disk. It is earmarked for very large applications with a single disk capacity of 6.8 gigabytes.

Section A.2.4 contains an analysis of the different types of writing methodologies that distinguish WORM disks.

A.2.3.3 Rewritable Digital Optical Disks

Most surviving methods of designing an "erasable" optical disk utilize magneto optics. Magneto optics use a vertical magnetization film for the recording medium and a laser for recording, replay, and erasure of the information instead of magnetic heads. Rewritable optical disks offer great potential to the market that needs erasability with massive storage capacity.

The process of writing to the magneto-optical disk can be described as the laser beam heating a pre-magnetized spot on the disk. When the spot is heated to the curie point, a small external magnetic field is introduced to the spot (see Figure A-15). The heating action

Magneto Optic Recording Principle

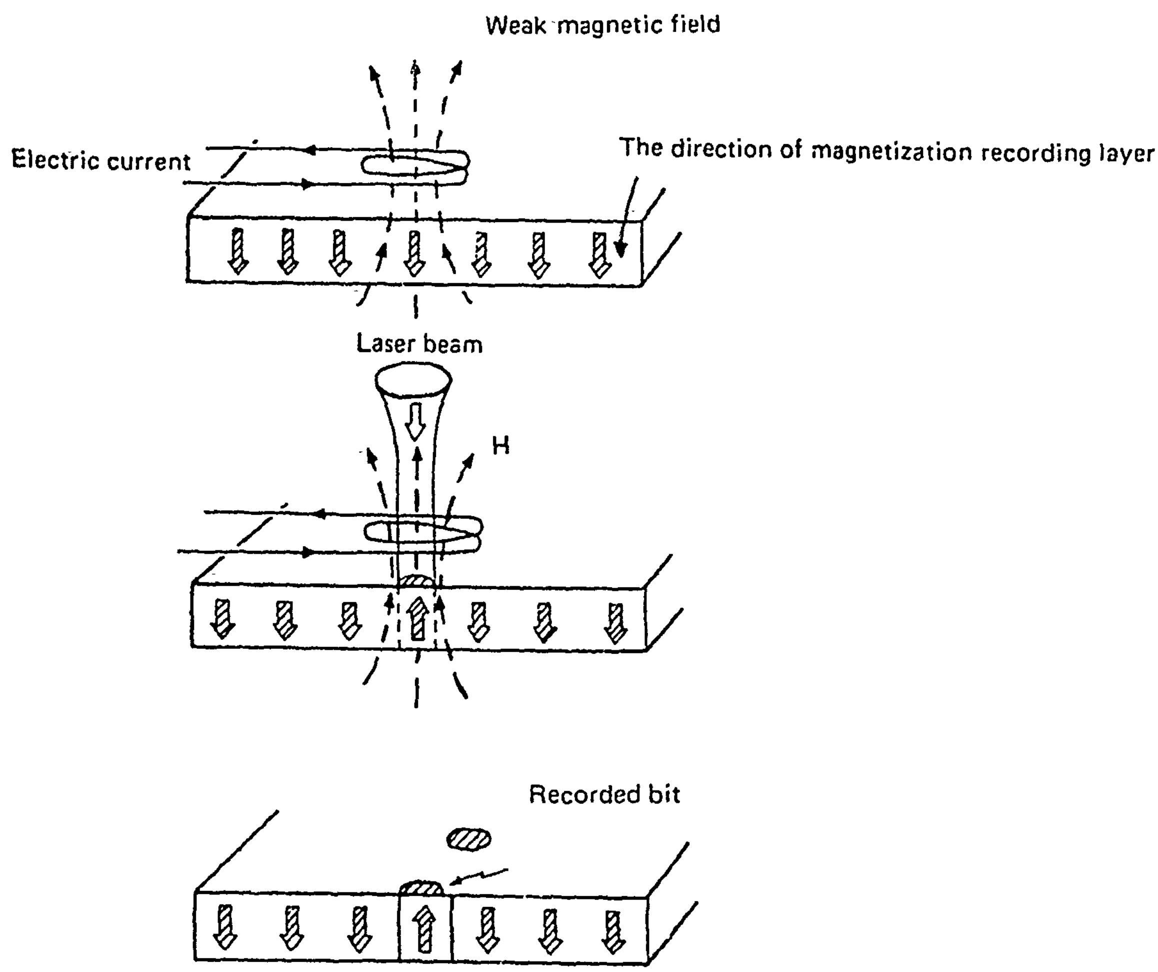


Figure A-15

enables the spot to reverse polarity and, thereby, carry the information. Reversing the process erases the spot.

Reading the disk is a separate process altogether. As the direction of magnetism of the spot is upward, its reflection polarization is rotated by a certain angle. Another angle is created with the downward direction of magnetism. This angle is interpreted by the sensors and converted to a binary signal (see Figure A-16).

Erasable or rewritable digital optical disks come in several different sizes for a variety of different purposes. Two-inch disks are not yet available commercially. They will probably have a storage capacity of from 20 to 50 megabytes. Three and one half inch disks are commercially available and have a capacity that ranges from 50 to 160 megabytes. Both of these very small disks are earmarked for applications that require a great deal of storage in a very compact space. Laptop computers offer a perfect application because of their need for data storage and a durable medium. Optical disks offer both attributes.

A.2.3.4 Digital Read-Only Optical Disks

Digital read only optical disks, Read Only Memory or ROM disks are very closely related to the videodiscs described in section A.2.3.1 on page 189. The main difference is in the type of data stored on the disk. Videodiscs store analog information and ROM optical disks store digital data. The majority of ROM disks are in a 4.72-inch (120mm) format with a user data capacity of around 550 megabytes. These disks are typically known as CD-ROM, Compact Disk Read Only Memory. A popular application for this medium is the compact audio disk or CD. This type of disk stores digital audio information and reproduces it with a fidelity and audio range unmatched by conventional [33 1/3 RPM] vinyl long-play records.

The methodologies and techniques involved in recording and reading information on ROM disks are similar to those processes described in section A.2.3.1 on videodiscs.

A.2.3.5 Digital Optical Tape

Digital optical tape is a relatively new product in the marketplace. It offers a write-once capability similar to WORM optical disks. Since it uses a tape format, it is, by definition, a sequential access medium. A large tape format wound on a 12-inch reel produces a very large surface area on which to store digital information. In fact, some manufacturers are advertising a one terabyte[78] capacity.

This type of medium is ideal for applications in which fast, random access to files is not required. Many system integrators are viewing this storage medium not for primary storage, but for image data backup where access time is not critical.

A.2.3.6 Digital Optical Cards

Digital optical cards are, simply put, credit cards with an optical instead of a magnetic recording strip. The technology is very similar to all the optical media discussed above. Currently, the card can be written to and read from using an inexpensive slotted reader-

[78] 1,000,000,000,000 (one trillion) characters of storage.

Magneto Optic Reading Principle

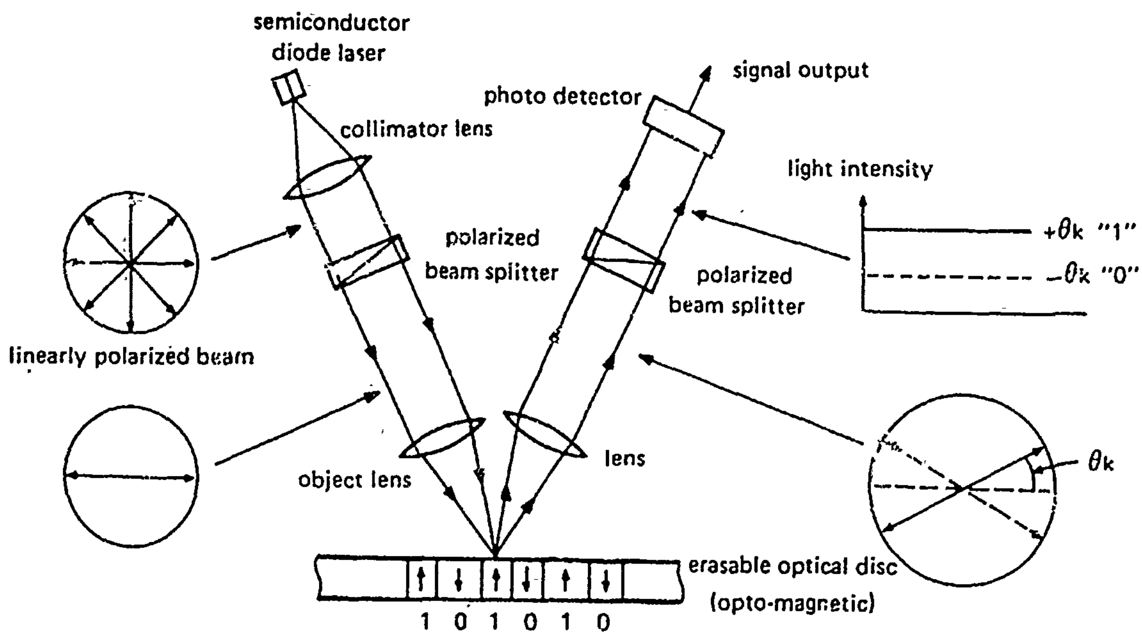


Figure A-16

writer much like those used for conventional credit cards. Capacities range from 2 to 200 megabytes.

These cards are being used in a variety of applications. They are useful for easy transference of data due to their small size. They can also be a good source of personal information such as one of the current applications at an insurance company. Individuals holding hospitalization policies would carry a card with their entire medical history.

A.2.4 Write-Once Disk Recording Methodologies

There have been three main categories of write-once recording methodologies for digital optical disks. They can be classified as Deformation, Phase Transitional, and Alloy methods. All three, which are shown in Figure A-17, will be discussed in this section.

It is important to consider the material that forms the disk itself. There are three main types of substances generally used for the disk substrates. Polycarbonate (PC) is very strong against impact, can withstand high temperatures, and is fairly resistent to moisture. Polymethyl methacrylate (PMMA) is the most transparent of all plastics but is susceptible to moisture absorption. Glass is the last substrate type. It is much heavier than plastic and is very expensive to polish to a perfect transparency. However, it is virtually unsusceptible to moisture and can withstand very high heat.

The general structure of a write once disk is very similar regardless of the recording methodology utilized. The disk substrate is either plastic or glass and forms the structure of the disk. Next, the reflective layer is the layer that carries the data and is usually formed by a metal alloy that is highly reflective, has a low melting point, and is not susceptible to oxidation. A protective layer and adhesive layer are next. A disk is really two disks that are glued together to form one, two-sided disk.

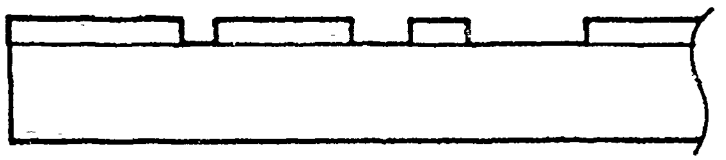
All WORM recording techniques use the same premiss that a change in reflectance at a particular spot on the disk recording layer is interpreted as the information. How that spot changes its reflective properties is the basis of the three different techniques mentioned above.

Deformation is when the laser heats the recording layer to a point that either raises a blister or bubble, or burns a pit or hole. The reading laser travels along the track at a certain rate looking for its reflection (see Figure A-18). When it encounters a hole or a blister, the beam is diffracted and the amount of reflection is changed (see Figure A-19). This change in amount of reflectivity carries the information. An area is required above the spot to allow room for the deformation to occur. The ablative material from the hole will spill over the top and reside along the rim. The blister will need room to grow without hitting the substrate.

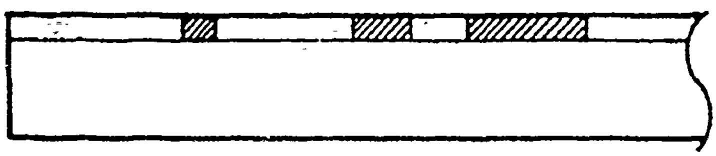
In order to accommodate these actions, an "air sandwich" is created which raises the substrate above the reflective layer. The cavity is filled with either an inert gas or a vacuum. The edges of the disk are sealed against moisture infiltration. One of the biggest threats to the longevity of an optical disk is the oxidation of the reflective layer. When even the most minute area of oxidation occurs, the reflectivity changes and, along with it, the information.

Phase transition changes the percentage of reflectivity by altering the structure of the reflective metal layer. Heat from the laser changes the normally amorphus metal to a crys-

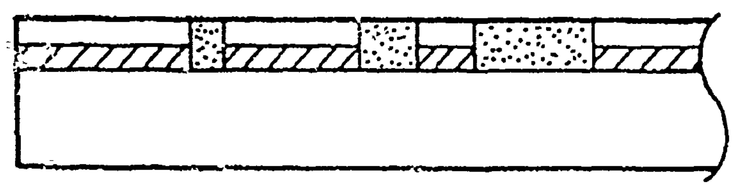
Write-Once Disk Writing Methods



Melt
(TOSHIBA, HITACHI)



Phase Transition
(PANASONIC)



Alloy
(SONY)

Figure A-17

Reflection of a Laser Beam

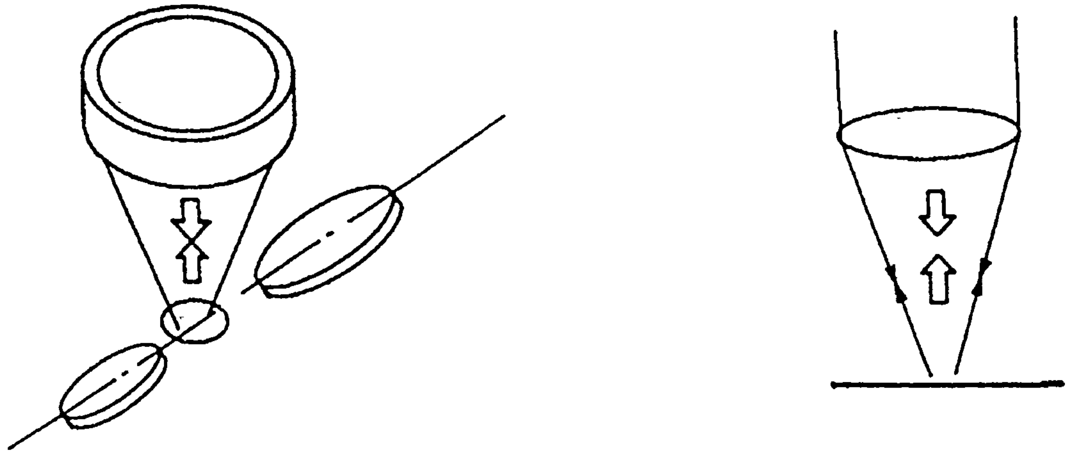


Figure A-18

Diffraction of a Laser Beam

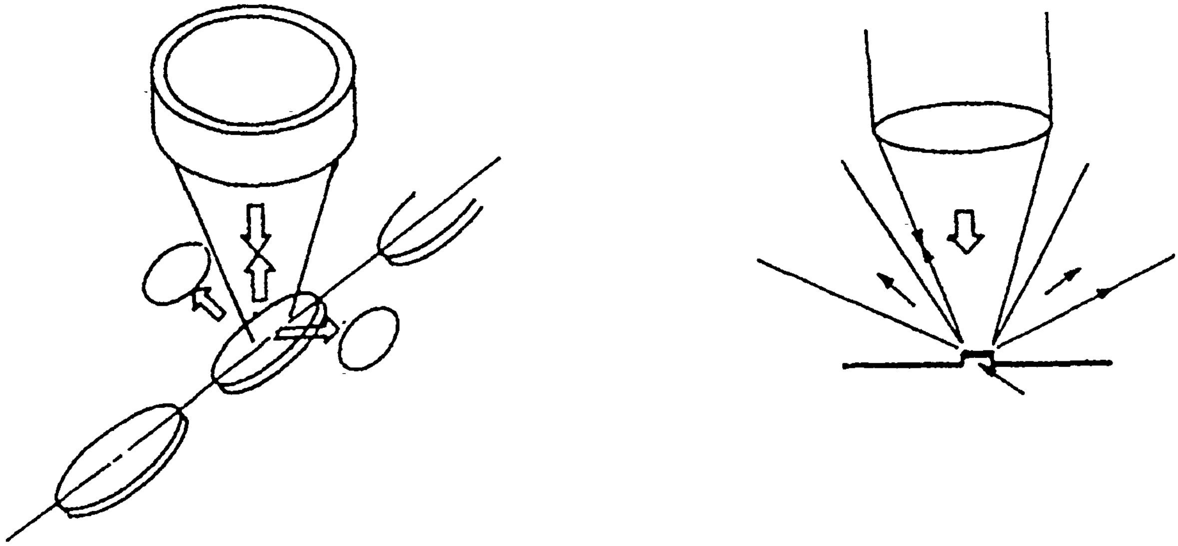


Figure A-19

talline compound with commensurate changes in reflectivity. The benefit of this process is that since there is no actual physical deformity taking place, no air sandwich is necessary. With no air sandwich, the possibility of eventual oxidation is greatly reduced. However, the amount of heat required to change the metal structure is difficult to regulate in such small amounts, resulting in a very delicate process.

The Alloy method is the latest technique employed to try to avoid the necessity of the air sandwich. In this case, two bi-metallic alloys are spun out onto the surface of the raw disk in two layers. The layer closest to the laser has a pre-determined reflective coefficient. In order to write information, a high-power laser melts the two alloys together at a particular spot creating a third alloy. The third alloy carries a very different reflective coefficient than the other alloy. The reading process is very similar to all the other optical disks. The low-power laser moves along the track and senses the change in reflectivity. That change carries the information.

A.2.5 Write-Once Recording Strategies

Since WORM disks are not rewritable, special care and planning must be invoked whenever they are to be used in a system. In applications that require document or file updating, some provision must be made for linkage of all the images within the file. This can be done either by saving room on the disk for later addition of more page images after the initial file has been written. Or, the linkage could be made logically by index pointers alone. That way, the disks could be loaded to their capacity without holding out space for late additions to files.

Another optical disk strategy worthy of mention is the separation of the reading and writing function. Writing the vast amounts of data that it takes to fill an optical disk can take quite a while. It may be prudent to separate the retrieval and writing functions by having more than one drive. This would enable uninterrupted retrieval without waiting for a write sequence.

A.2.6 Automated Retrieval Devices

When any digital image system using optical disks as the primary storage medium is large enough to have multiple disks, it is time to consider an automated retrieval system. There are many reasons that an automated retrieval system (jukebox) is justifiable in a digital image system. Without one, the user of the system would have to load his own disks. This may not be much of a problem in a small system with only one or two workstations. But, in large systems, an operator would have to be on call at all times that the system was up in order to retrieve disks from a shelf, put them in the drive, and replace them. This is a slow process, but may be cost effective in some applications where costs are more important than performance.

Another alternative would be to have every optical disk resident in its own drive. This would eliminate any need for loading and would provide the fastest access, but it would be very expensive.

Most digital image systems, using optical disks as their primary storage, use a jukebox in which to store and retrieve their disks. There are commercial jukeboxes for virtually all popular types of disks. A jukebox is a device that will store a number of disks, robotically retrieve them, and place them in a read/write drive. Typically, the jukebox would be

completely integrated into the system. During an image retrieval, it should be transparent to the user where the image is physically stored. The computer and indexing software would identify the location of the image file and control the actions of the jukebox.

There are several important considerations to make regarding jukebox design and implementation. The first is the number of drives relative to the number of disks stored. If a jukebox only has one drive for fifty disks, access contention for that drive will be very slow. There could also be contention for the robotic arm or picker to transport the disk to the drive and back. Large systems that require frequent requests should consider several smaller jukeboxes with multiple pickers and drives in order to limit the queue up time to get to the drive.

A.2.7 Optical Media Longevity and Stability

Optical media have not been around as long as some other types of storage media. There are some early disks that have survived as long as ten years. In most cases, however, we must rely on accelerated life testing done in labs to predict the longevity of the media. All manufacturers guarantee at least a ten-year life of the disk. Some will guarantee one hundred years. All of this discussion of the absolute longevity of the media may have missed the point.

It seems that, in the case of human-readable media such as microfilm, the life of the piece of film is relevant since it would be possible to read the film given sufficient light and a magnifying device of some sort. Optical disks, however, are not human readable. In fact, they require a fairly sophisticated computer system to read the disk and interpret the binary 1's and 0's to form something that can be human readable. Therefore, the life cycle of the media cannot be considered separately from that of the system required to retrieve it.

There are other factors to consider, as well, in the question of media longevity. The recorded data on the optical disk is digital. Since there are reliable techniques to transfer digital data between different media types without any loss of data or generation, the data could migrate to another medium at the end of useful life of the optical disk. It then becomes a question of cost beneficiality. The percentages of storage currently being used indicate that the capacities of optical disks, as currently known, will continue to increase within the current price structure. That means that the cost per bit will continue to decrease making transference a practical matter. Migration of digital data would be relatively straightforward and not involve any rescanning or paper handling of any kind.

The National Archives is currently underwriting a laboratory project at the National Institute of Standards and Technology to conduct optical disk longevity tests independently of vendors. They will try to develop a standard testing methodology which could be used to determine the useful life of optical media.

A.2.8 Legality of Digital Images From Optical Disks

The question of the legality of digital images from optical disks is important to consider relative to the disposition of the original documents after conversion. If they are to be destroyed or moved to a virtually inaccessible location, the operation will have to rely on the digital image to stand on its own in all legal forms and situations.

Legal opinions, thus far, have been narrowly structured to encompass only the confines of the particular case or application. There have not been any court opportunities to create a test case that would offer adequate precedent for a wide range of application areas. Failing that, most user applications are simply adding "digital image from optical disk" to any law or regulation that includes the legality of microform.

Given similar security safeguards found in current computer systems, a digital image system that uses WORM disks should not have any problem in qualifying for the same treatment as an analog image stored on microform. Neither image can be easily changed and both offer a true and accurate representation of the original.

Since digital image technology does offer such capability in the area of image manipulation, and since scanning does not capture all of the original document, it is important to maintain a complete audit trail of all actions taken. Evidence introduced into court will always need to be the best evidence available. If the only version of the original document is a print from a digital image system stored on optical disk (i.e., the original no longer exists), it must be created in the normal course of operations. That is, the originals must normally be destroyed after scanning and not just in this single case so that the best evidence would be the digital image.

Local, state and federal agencies will probably begin to certify their own operations and certify that the image in question is a fair and accurate representation of the original document as is currently done with photocopies and microform prints.

A.2.9 Standardization

There are different types of standards that are important in the world of digital image and optical disk technologies. They can be classified as *de facto* and government regulations. The *de facto* standards take several forms themselves. There are the industry standards adopted by popular demand. There are sanctioned standards set up by accredited standards bodies such as the American National Standards Institute (ANSI), the IEEE, and the International Standards Organization (ISO). There are also non-sanctioned bodies such as the Association for Information and Image Management (AIIM) that contribute with their own standards committees.

The federal government also issues a variety of different regulations that could cover all aspects of image conversion, storage, retrieval, and permanent disposition.

There are several projects ongoing within the ANSI standards group known as X3B11 on digital optical disk technology. The group has on its agenda 5.25, 12 and 14-inch WORM disks and 3.5 and 5.25-inch erasable disks. At this time, there has been no final standard issued that is comprehensive enough to enable the user to make procurement decisions based on its implementation. It will probably be at least a year before definitive standards are completed and published.

APPENDIX B

DETAILED ODISS SUBSYSTEM DESCRIPTIONS

APPENDIX B. DETAILED ODISS SUBSYSTEM DESCRIPTIONS

B.1 Basic System Concept

The Optical Digital Image Storage System was designed to be a useful laboratory to test various aspects of both digital image and optical disk technologies. As a research lab, its design must lend itself to the flexibility required for testing a variety of materials and operations. This section describes the overall design and work process flow. The following section describes the general hardware configuration and operational process. For more detailed hardware descriptions, refer to sections B.2 through B.12.

B.1.1 Configurational Scheme

The system design is divided into three main subsystems: conversion, storage, and retrieval. The conversion subsystem contains the document preparation, scanning, indexing, quality control and rescanning functions. The hardware required to accomplish this operation consists of a TDC model 4200 high speed, two-sided paper scanner (modified by Photomatrix Corporation) for primary input conversion. Index and quality control workstations are based on Sperry IT personal computers with Discorp high resolution screens and decompressor boards. The rescan station utilizes a Ricoh RS400 scanner with an IPT enhancement board powered by a Sperry IT personal computer.

The storage subsystem takes the prepared images and index data and stores them for subsequent retrieval. Sony 12-inch, Write Once Read Many (WORM), digital, optical disks are used for long-term image storage. These disks are stored in a Sony Autochanger (jukebox). Index data is maintained on magnetic disk storage.

The retrieval subsystem uses Sperry IT personal computers and Discorp high resolution screens and decompression boards. Ricoh laser printers provide hardcopy output. There are two remote image workstations. One is located in a staff office area, and the other with a printer is located in the public reference room. A remote (index only) workstation and dot matrix printer is placed in the Tennessee State Archives in Nashville, Tennessee.

B.1.2 Capture and Retrieval Process

The operation of the ODISS test configuration is generally described in this section and in Figure B-1. The ODISS system operation is shown in Figure B-2 and Figure B-3.

The object of the capture subsystem is to scan an original document and create an electronic image, using a high speed process, that adequately represents an original paper document. This image is indexed for later reference, reviewed for quality and index accuracy and prepared for long-term storage. During the conversion processing, the image and index data reside as electronic files on a magnetic disk buffer. In the few instances that the high speed scanning fails to produce a facsimile image of sufficient quality, the scanning process is repeated, using another scanner that is capable of a higher degree of operator interaction to produce superior image quality, but at a lower throughput rate.

The blocks of completed files are finally ready for the process of the long-term storage of the images and index data. The index data is maintained on magnetic disks while the images are written to digital optical disks. In order to accomplish the transfer and at the same time

System Block Diagram

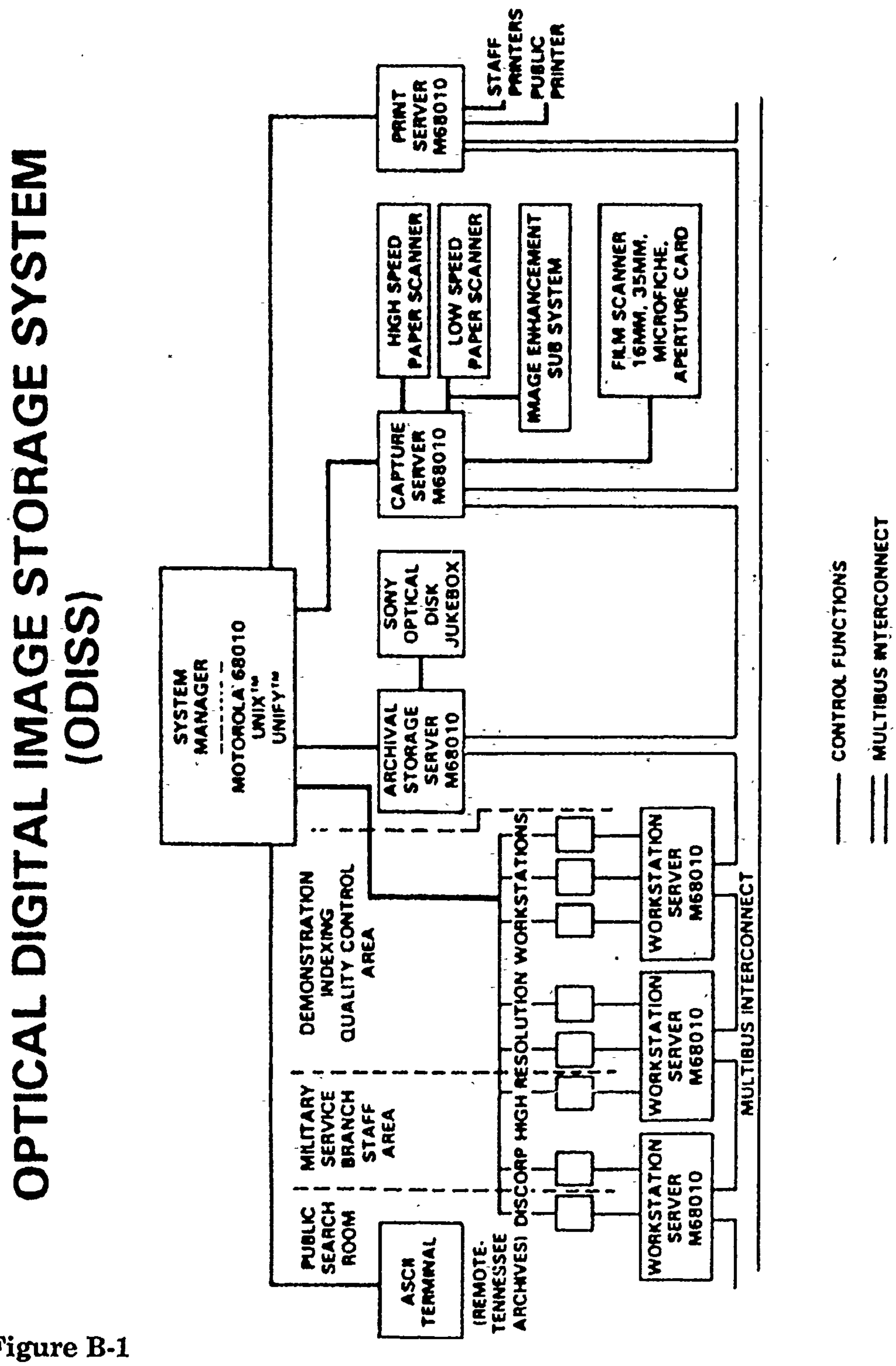


Figure B-1

Capture and Storage Subsystems

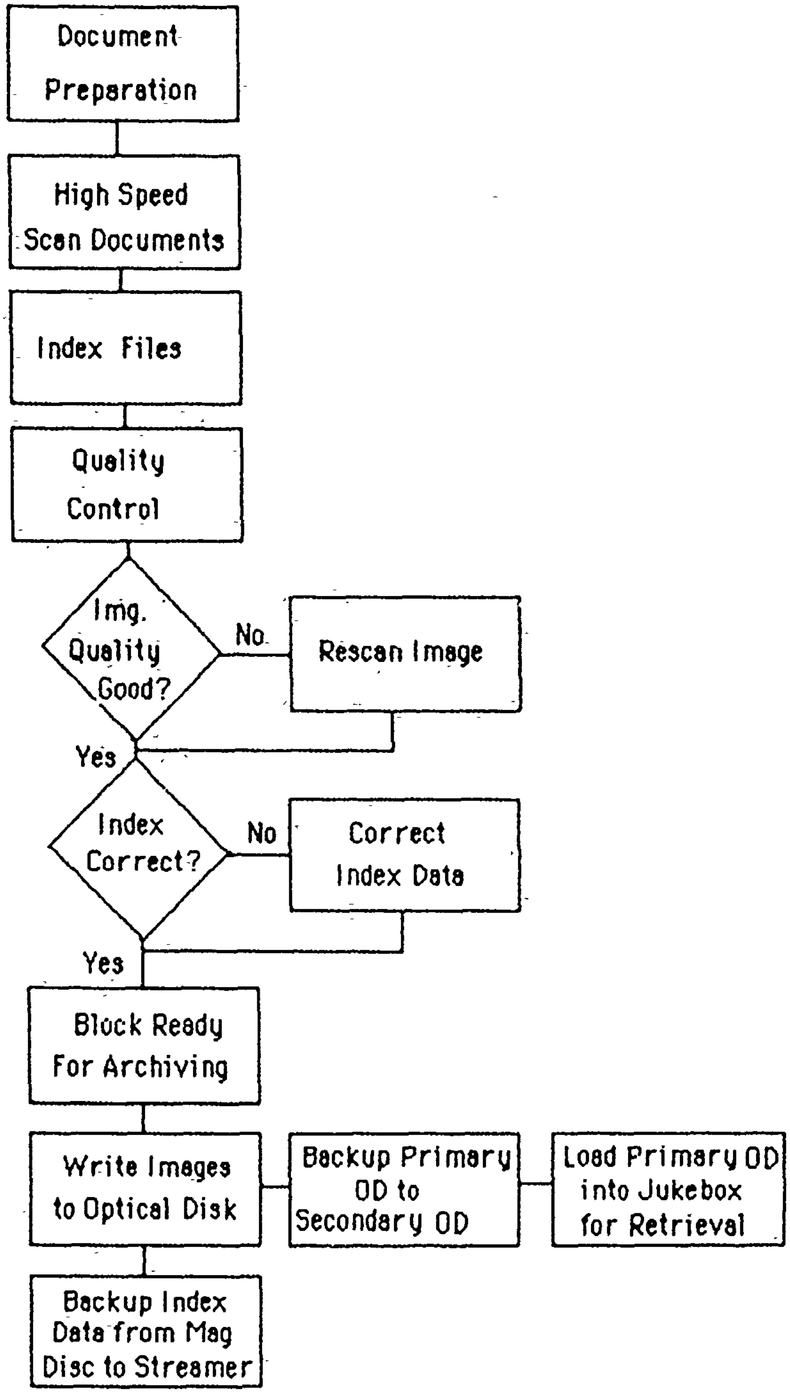


Figure B-2

Retrieval Operation

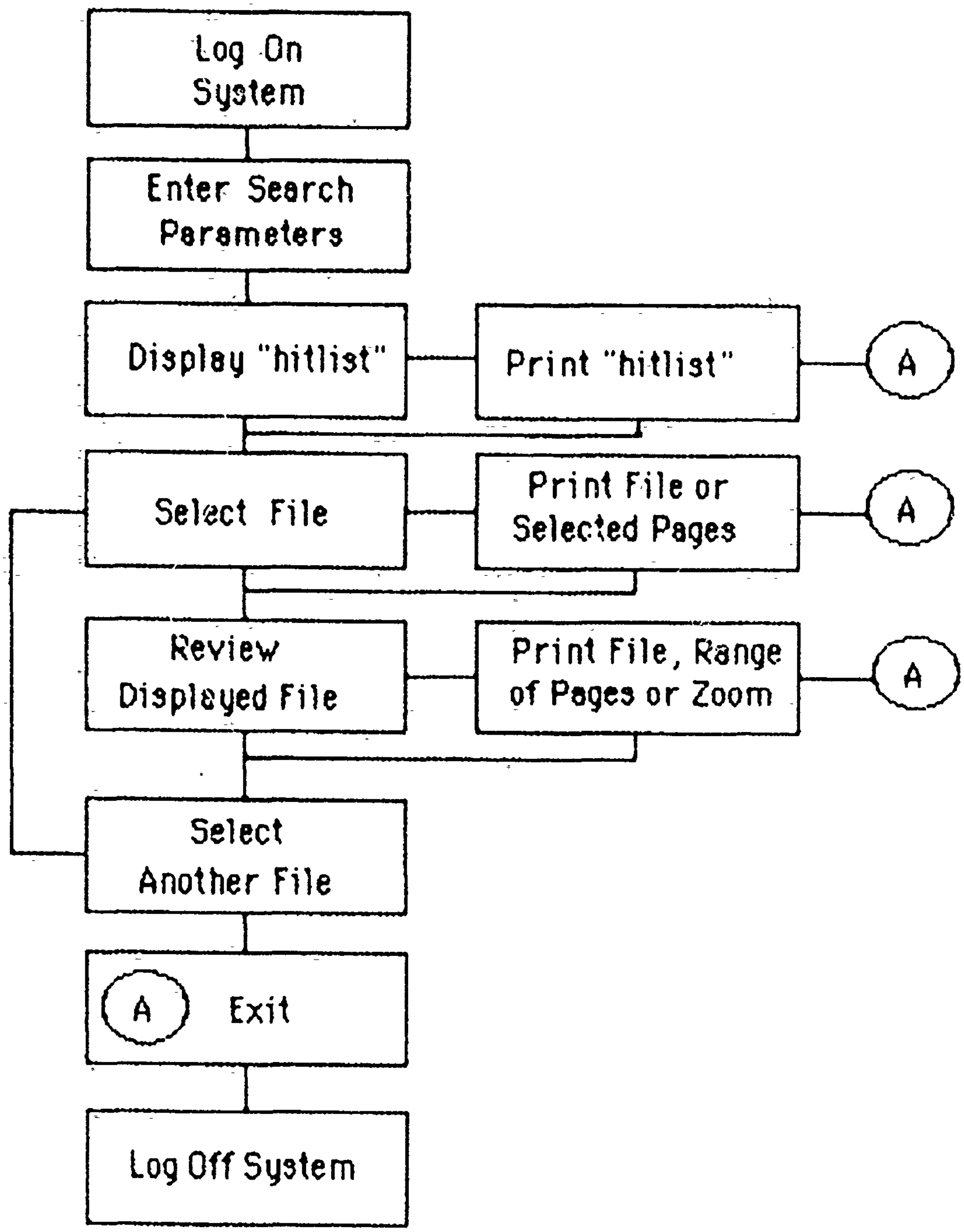


Figure B-3

reduce "out-of-file" conditions, several operational steps must be followed. The images are written to side "A" [of two sides] of the first optical disk. Once side A is full, it is "backed-up"[79] or copied to disk two, side A by using an additional drive. When completed, disk one is flipped over in order to allow recording on side B. Disk two remains in the back-up drive so that access to those files (side A) is possible while the primary or disk one is on side B. Once disk one, side B is full, it too is copied to disk two. Disk two (the back-up disk) is stored elsewhere and the primary disk (disk one) is loaded into the jukebox for retrieval. Using these procedures, access to file images is possible anytime after the file is indexed.

B.2 Digital Image Scanning

There are four types of scanners in the ODISS configuration. Three of these scanners are for documents, while the fourth is used for scanning different formats of film. All scanners perform the digitizing of image information that forms the basis of the NARA image data archive.

B.2.1 High Speed Paper Scanner

The Photomatrix high speed scanner is designed to scan both sides of documents creating image data from them at a rate in excess of 20 documents per minute. The high speed scanner is made up of two components, each contained in its own separate enclosure. The first is the scanner transport unit, consisting of a Terminal Data Corporation (TDC) model 4200 paper feed mechanism and scanning electronics. The second is a set of electronic components that control the scanner and handle the data that the scanner provides. This second piece of equipment is referred to as the high speed scanner electronics. Also included with the scanner is the high speed scanner monitor, which displays images as they are captured by the document scanner. Table B-1 gives the scanner's technical specifications.

The main components of the scanner are two sets of vacuum panels, belt mechanisms to feed the documents, charged coupled device (CCD)[80] arrays for scanning, and florescent lights to illuminate the documents being scanned. Two sets of these components are needed to scan both sides of the document quickly. Electronics to convert the CCD output into digital image data are also provided within the scanner. There are also several large motors to run vacuum pumps and the document feed mechanisms.

The high speed scanner operates on the principle of moving the document in the subscan direction while scanning in the main scan direction with a linear CCD array in order to scan the complete document. As the document passes through the field of view of the CCD array, light is reflected off the document and onto the CCD. Analog data from the CCD are converted to digital data and sent to the high speed scanner electronics. The signal is processed by the scanner electronics, and, using a number of different techniques, the result is displayed on a high-resolution monitor at the high speed scanner workstation. Image data are then compressed prior to transmission to the rest of the ODISS system.

---

[79] To create a duplicate copy for security or disaster recovery purposes

[80] A CCD is a device that converts light into an electronic signal.

High Speed Scanner Specifications	
Type	Belt Transport
Scan Method	3,456 CCD element array
Number of Sides	2
Page Size	8.5 X 14 inches maximum
Density	200 DPI
Scan Rate	20 to 26 PPM for 8.5 X 11 two-sided pages depending upon compression factor attained. Smaller documents are proportionally faster. Work to develop a faster rate for single image pages is proceeding.
Processing	5 X 5 convolution filter
Data Compression	CCITT Group III
Video I/O	RS-422
Channels	1 1-bit
Clock	10 MHz
Control I/O	RS-422
Rate	9600 Baud
Mode	Full Duplex
Monitor	19" black and white

Table B-1

The two sides are scanned by first holding the document against the bottom feed belt, which has pinholes through it to allow the vacuum mechanism below it to hold the document to the belt (see Figure B-4). The belt transports the document through the field of view of the reverse scan CCD. From the lower belt the document traverses to the upper belt which moves it through the field of view of the obverse scan CCD before dropping it into a hopper.

The scanner is controlled by a Heurikon MLZ-93 microcomputer, a single board computer based on the Z80 microprocessor. The board has 128K of RAM, 32K PROM memory, two RS-232 serial interfaces, and two interfaces to the scanner electronics multibus. Also on the multibus is a memory expansion module providing 2 MB of buffer memory for use by the MLZ-93 microcomputer.

Docuscan DS-4000 High Speed Transport

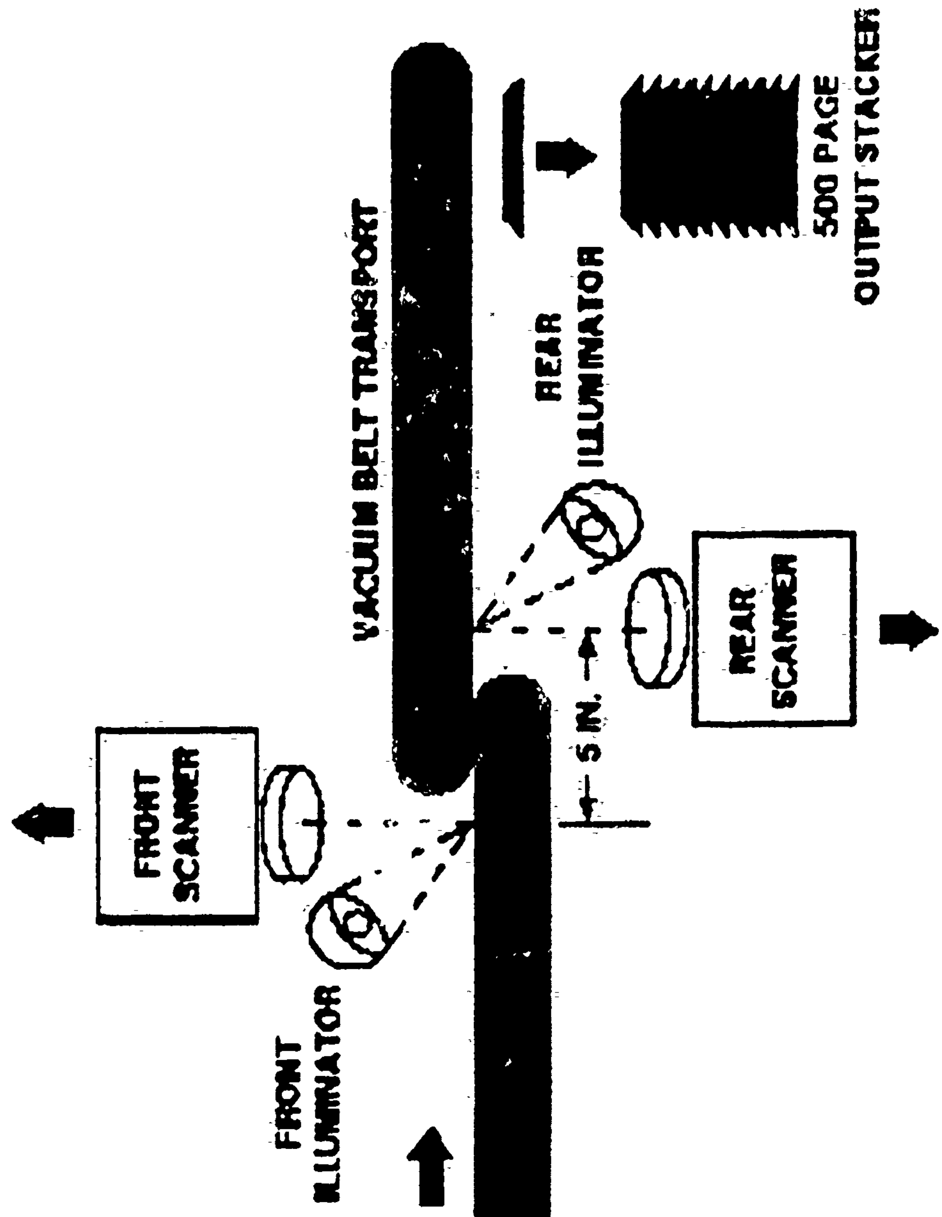


Figure B-4

In the scanner electronics cabinet, the output from each of the scanner's CCDs is stored in a five-line-deep buffer that is used for image processing. The processing technique uses a five-by-five matrix to perform weighted sum convolution filtering[81] to determine the value of the center pixel[82] of the matrix. The convolution is performed by high speed adder and multiplier chips that output a 12-bit value equivalent to the two's complement of the matrix result. The result is then thresholded against three different levels: the light threshold, the dark threshold, and a third threshold to determine the value of all pixels that fall between these two levels. The resultant two-tone data are then compressed[83] by a Kofax 9000 compression expansion board and output to the host[84].

The electronics within the high speed scanner are sufficiently powerful to process and compress four pages of captured data. The image data are kept in one of four Kofax compressor boards, the last stage before the image is output, while it waits for a logical connection that will allow the data transfer.

The scanner electronics are interfaced using one RS-232 and one RS-422 cable. The RS-232 cable is used for control of the scanner and status information to the Hk68/M10, and the RS-422 cable is used for transmitting compressed image data. These data are sent to a special Unisys interface on the HK68/M10. The scanner is directly controlled by a small panel on the control board of the scanner. The console features an LED[85] message bar, a 6-key selection keypad, and a 9-key control keypad. The message bar provides feedback to the operator to help him operate the scanner, and to warn him of any errors or malfunctions. Keys on the 6-key pad are: three size keys to choose a width for the document being scanned, a one sided/two sided scan toggle key, and keys to reset the scanner and take the scanner on/off line. Three keys on the 9-key pad are for controlling the thresholding of the image by the image processing electronics. The other six keys in this keypad are for providing information to the computer for controlling blocks and files in file-oriented operation.

Communications between the components of the high speed scanner system are carried out on a number of signal cables running between the high speed scanner electronics box and the high speed document transport. Six cables pass from the transport to the electronics box. Two of these cables, ADC control and ADC clock are for controlling the electronics within the transport. Two other lines carry data from the two CCD arrays to the electronics box. Another line carries information from the transport to the electronics box. The last line to pass between these two units is the control line to the transport mechanism. One line is a 25-pin female socket that supports an interface to a stand-alone terminal that can be used when the scanner is in local mode. The display port is a DB-25 pin connector that transfers uncompressed image data to the image monitor for display.

---

[81] A specific image enhancement technique or algorithm

[82] Abbreviation of "picture element" or one of millions of small dots that make up the image

[83] Method by which redundant data streams are reduced to a much smaller sizes

[84] The controlling computer

[85] Light emitting diode

The high speed scanner interface consists of a printed circuit board mounted within the Core enclosure and connected to Heurikon board VRTX0. The interface contains FIFO[86] buffers, driver/receivers, and logic gates. Its purpose is to interface data, timing, and control signals between the scanner and the HK68/M1O VRTXO. The interface accepts compressed image data from the high speed scanner in the form of RS-422 differential signals and at a 10 Mhz rate. These data are buffered in FIFOs and passed to the iSBX circuitry on the Heurikon single board computer VRTX0. The iSBX circuit converts incoming serial data to parallel data. These data are transferred to on-board memory via single address DMAC data transfers. These compressed image data are then passed to the other Heurikon single board computer in the Capture Subsystem, VRTXI, for temporary disk storage.

The high speed scanner monitor is based on the Unisys PC/IT and the Discorp workstation. The monitor simply displays image information captured by the high speed scanner. The interface that transmits image data to the monitor is at the rear panel of the high speed scanner electronics. These data are provided to the image monitor in uncompressed form, and is displayed immediately upon receipt.

The purpose of the monitor is to provide immediate feedback for the user of the high speed scanner, i.e, allow the user to verify whether or not the scanner is operational and calibrated. The monitor supports basic image manipulation functions. The display can be toggled between two different display modes. The first mode is to display a subsampled set of data at 150 dpi[87] so that an entire image can be displayed on the screen. The other mode is to display each pixel of captured information at 150.dpi. This has the effect of enlarging the image by 30%, and shows the data captured by each element of the high speed scanner CCD array.

High Speed Scanner Operation

The Optical Digital Image Storage System (ODISS) high speed document scanner is the large volume entry for page images. The pages entered through the scanner are converted into compressed images and transmitted to the Capture Storage Element (CSE) where they are stored on magnetic disk for later examination. Once the quality assurance analysts are satisfied that the images are the best ones obtainable, they are indexed, tabulated, and stored on optical disk for long term storage.

The high speed document scanner is controlled by two ODISS elements: the high speed service element (HSE) and the high speed capture element (HCE). These computer program elements reside on different computers that are each integral parts of ODISS. They are connected to each other and to the system manager element and capture storage element. The HCE is a background element that controls the scanner operation and light panel and sends the images to the CSE. The HSE program interacts with the HCE, handles the scanner control terminal, and interfaces with the ODISS System Manager. These programs are identified for instructional purposes. Their existence should be transparent to any scanner operator.

---

[86] First in, first out

[87] Dots Per Inch is a method of defining image resolution or definition.

High Speed Scanner Control Terminal

There are two terminals near the scanner. One is the operational control terminal; the other is the diagnostic terminal. The operational control terminal (see Figure B-5) contains a four section picture: a large section across the top that is used to display block and file numbers; a narrow window across the middle of the screen that contains status messages; a summary in the lower left that describes files; and, the lower right section shows the currently active menu. The diagnostic terminal contains information about the scanner: status summary, help menu, or results from a diagnostic action.

Normal High Speed Scanner Operation

Scanner operation was developed on the "block of files" concept. Each block contains one or more files of documents that are to be entered and controlled by block numbers and file numbers within each block. When the upper window of the control terminal shows a block number, that block number is ready for use. If the large number is displayed in "solid" numerals, the block is open. Until it is open, the number is displayed as "non-solid" numerals. The display also includes words indicating the status: block open or block closed. The file number is handled in the same manner. Normal procedure is to collect the documents to be entered, obtain a block number, determine how many files there are in the block, and request the use of the block number by entering the number at the cursor. If the displayed block number does not agree with the required number, enter the number desired. If it is available, the number is displayed. If it is not usable, the next available block number can be requested by depressing the function key corresponding to "request next block" in the block menu.

Once the initial block number is located and displayed, block starting, file starting, changes to next block or file, and ending actions are handled at the scanner console. Alternate block and file start and end buttons are provided at the control terminal.

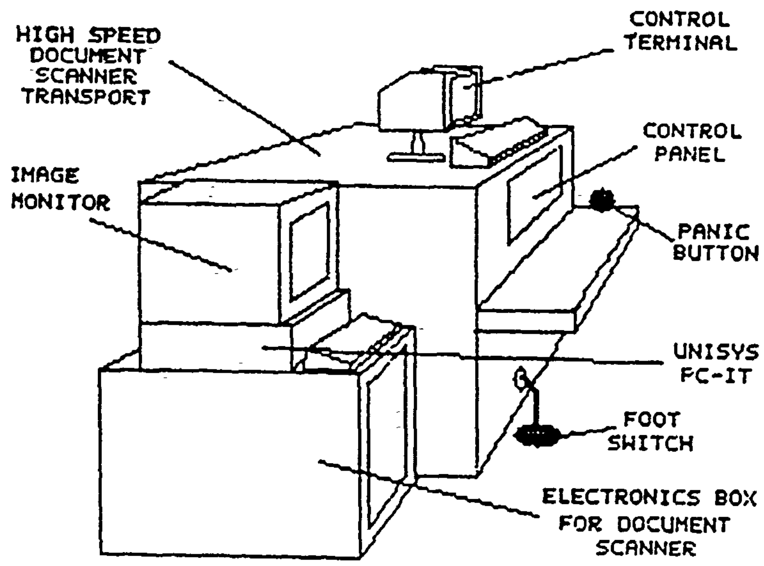
The block and file menus show the currently permissible actions next to the function key identifiers that when depressed invoke the desired actions. The function keys are the ten keys at the left edge of the terminal's keyboard, labeled F1 through F10. The menus include these key identifiers.

Scanner Control Console Operation

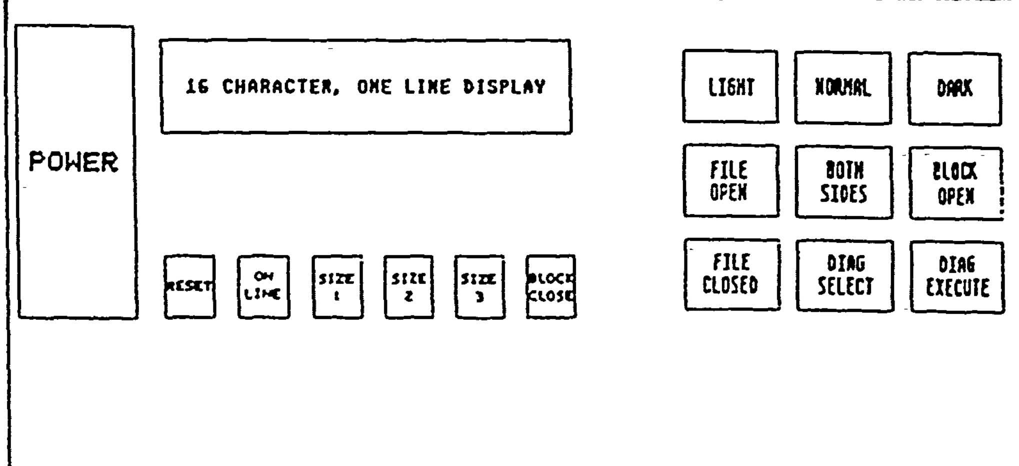
The high speed scanner console or operator's panel includes several lights and light/buttons that are utilized to indicate and control system actions. For example, to enter a block of images, the operator pushes the block open indicator button. The button lights up immediately, and remains lighted until the first document is read or until the system is reset.

When the scanner is properly adjusted for the documents to be read, operation is simple. The user need only be aware of the block and file status lights and the scanner display panel. If a document is misfed, or one of the scanner interlocks is active, the operator must use one of the features discussed in the following paragraphs.

Scanner Button/Indicator Panel



High Speed Scanner Position



Scanner Button/Indicator Panel

Figure B-5

Normal operation of the scanner begins at the operation control terminal. Figure B-6 through Figure B-9 are representative displays at the terminal during an opening of block 19 and creation of three files. The terminal display should show a block number in a non-solid form. The next available block number is fetched from the System Manager. When it is received, a crosshatched number appears under the block number field. When the desired or next block number is ready, i.e., a crosshatched number shown in the block number field of the display (Figure B-6).

To start the scanning process, depress the BLOCK OPEN button. While the system is performing this block and file opening process, make sure that the documents to be entered into the first file of the block are in order and accessible. When the BLOCK OPEN and FILE OPEN buttons are on, the system is ready for scanning (Figure B-7). These lights go off after the first document is read. This helps the operator to remember the current status. When the block and/or the file is closed, the BLOCK CLOSE and FILE CLOSE lights turn on. Documents are placed on the scanner feeder belts face down with the top of the document toward the machine. The document moves in a straight line and falls, face down, in the stacker tray. Hence, the documents are in the entry order when removed from the stacker. The feeder belts activate when a document is placed against the alignment bar at the red line near the entry. The suggested procedure is to place the document face down about one inch left of the alignment bar, slide the document forward until it is near the entry line, and then gently push the document to the right until the feeder activates. The vacuum system activates first, then the belt drive starts. When the feeder belt stops, the next document may be loaded.

When all documents for a file are entered and another file is ready for entry, depress the FILE OPEN button. When the FILE OPEN and BLOCK OPEN indicators light, the system is again ready for capture (Figure B-8). This procedure is continued until the last file of the block is completed. If there is another block of documents ready for entry, depress the BLOCK OPEN button. The current open file and block are closed, and new ones are opened, as before.

When the last file to be entered is completed, the BLOCK CLOSE button is depressed. The panel will indicate when the block is effectively closed (Figure B-9).

Image Monitor Operation

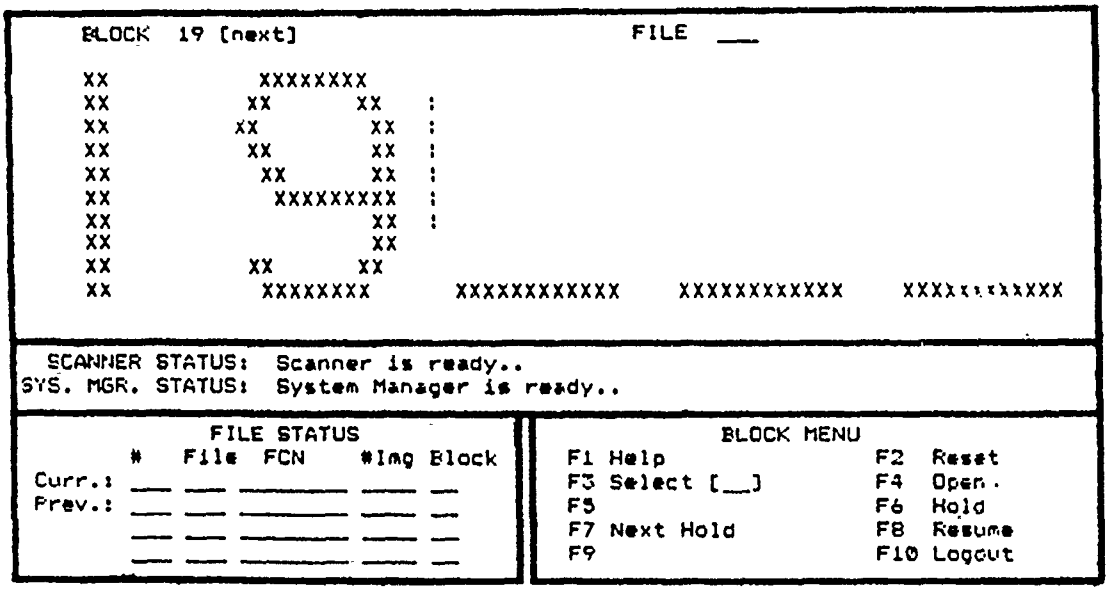
The high speed scanner image monitor is based on the Unisys PC/IT[88] and the Unisys-2000 workstation. The monitor simply displays image information captured by the high speed scanner. The interface that transmits image data to the monitor is at the rear panel of the high speed scanner electronics. These data are provided to the image monitor in uncompressed form, and are displayed immediately upon receipt.

The purpose of the monitor is to provide immediate feedback for the operator of the high speed scanner: i.e., allow the operator to verify whether or not the scanner is operational and calibrated. The monitor supports basic image manipulation functions. The display can be toggled between two different display modes. The first mode is to display a subsampled set of data at 150 dpi so that an entire image can be displayed on the screen. The other mode is to display each pixel of captured information at 200 dpi. This has the effect of enlarging the image by 30%, and shows the data captured by each element (top/bottom) of the scanner.

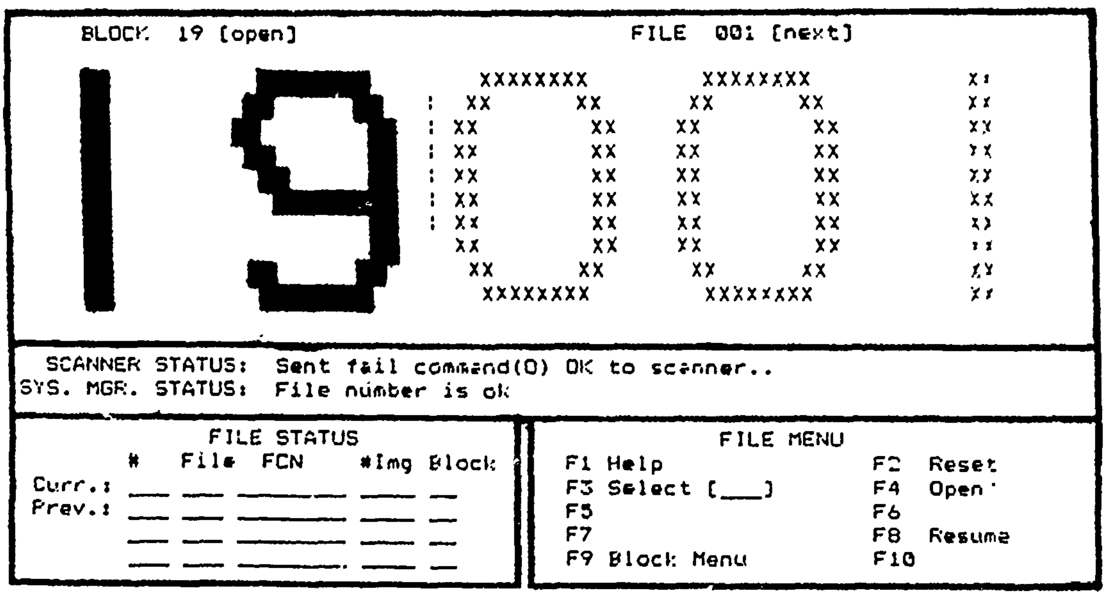
---

[88] Unisys version of a 80286-based, IBM-compatible, personal microcomputer

Operational Control Terminal; Awaiting Block Open



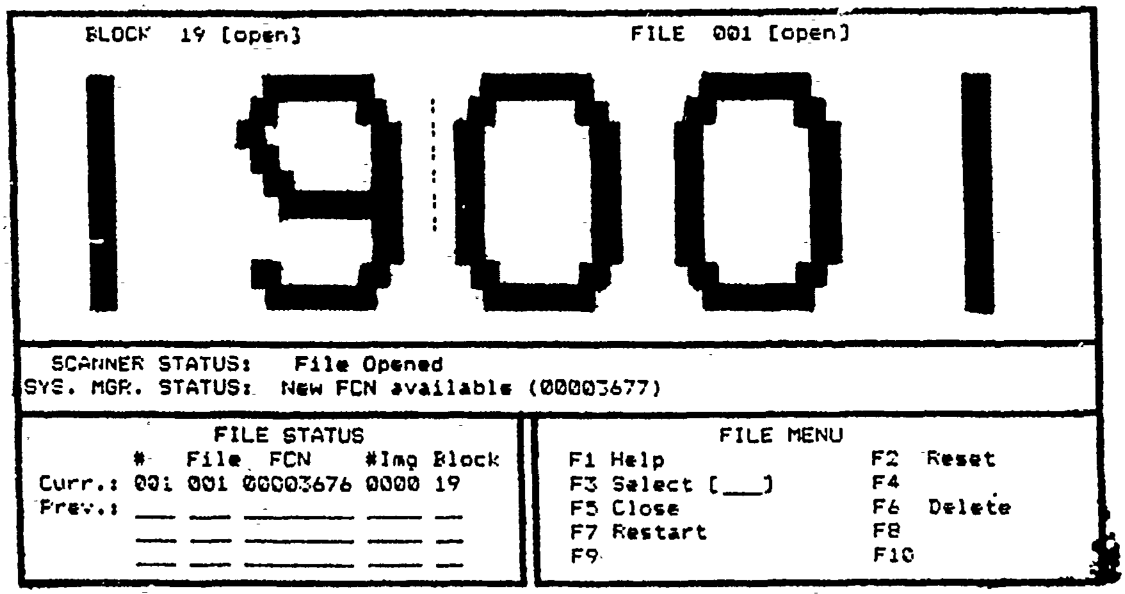
"Scanner is Ready"



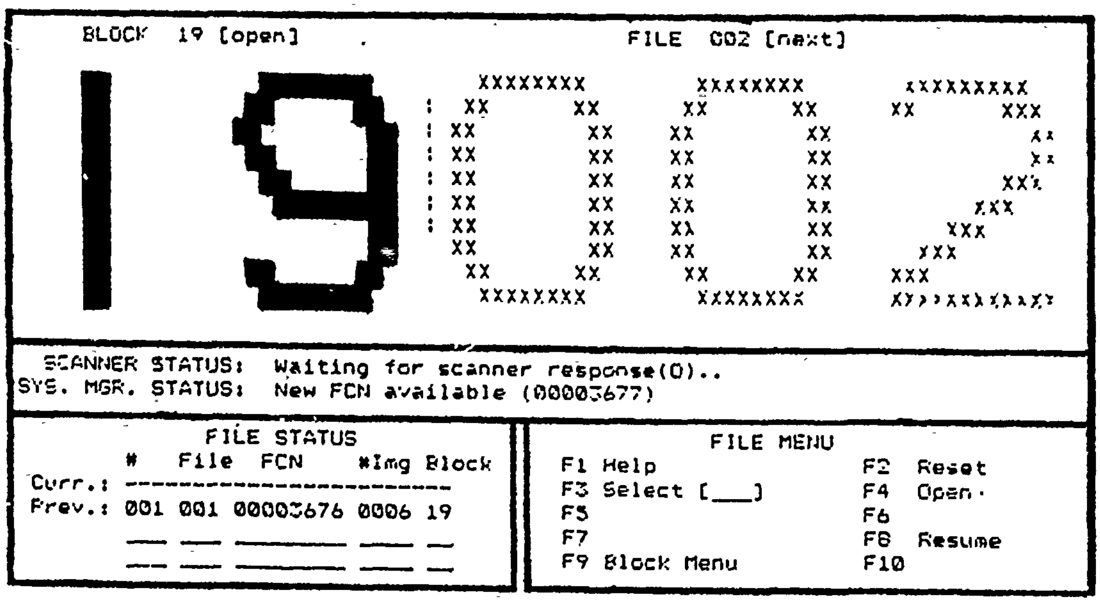
"File Number is OK"

Figure B-6

Operational Control Terminal; Ready for Scanning



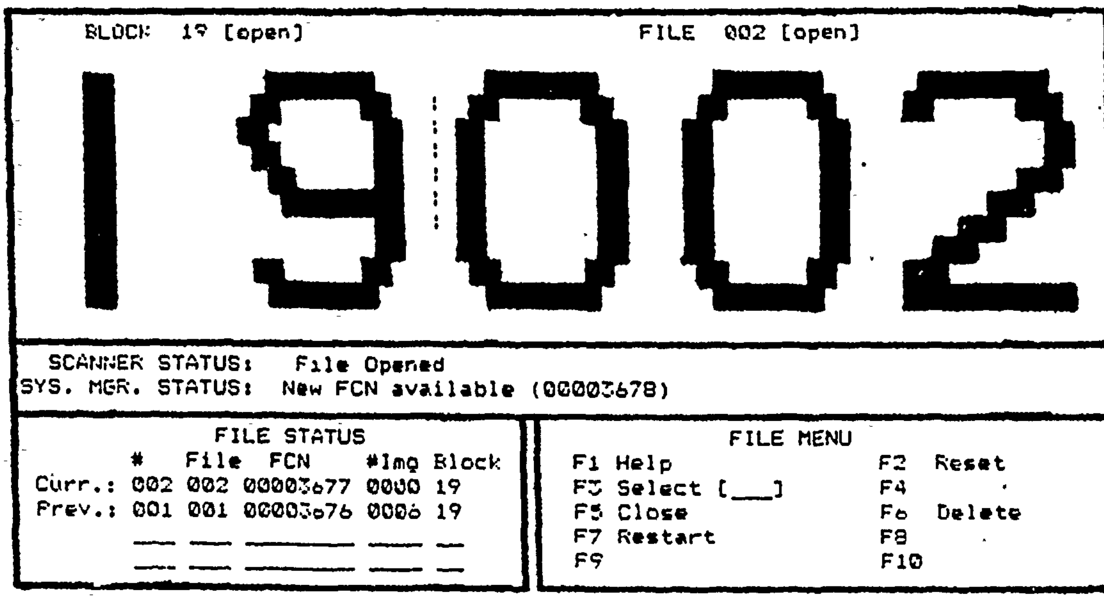
"File 001 Opened"



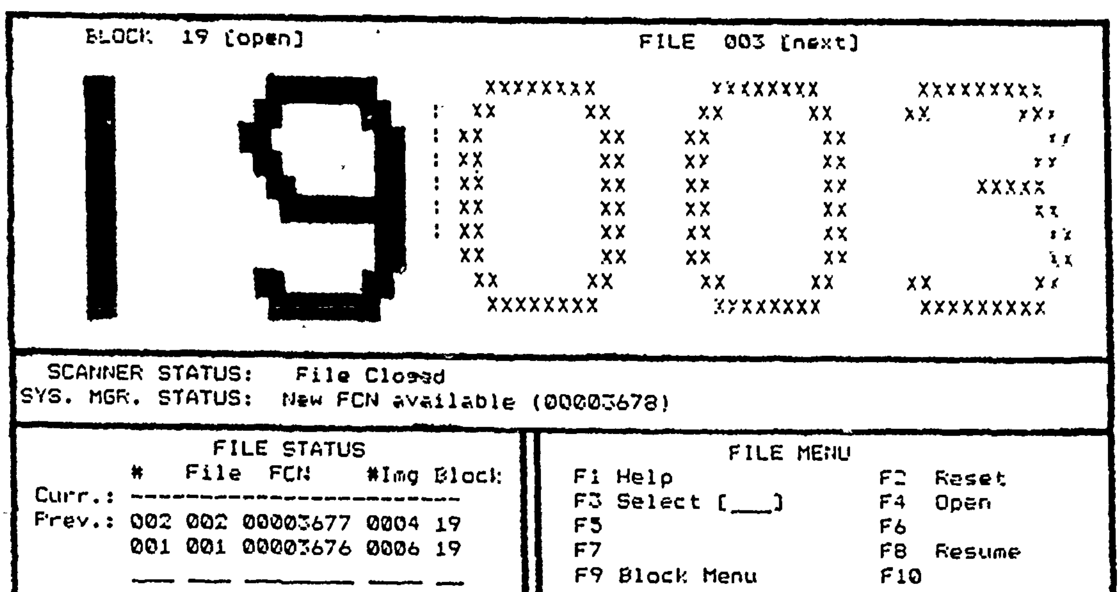
"New FCN Available"

Figure B-7

Operational Control Terminal; Next File Ready



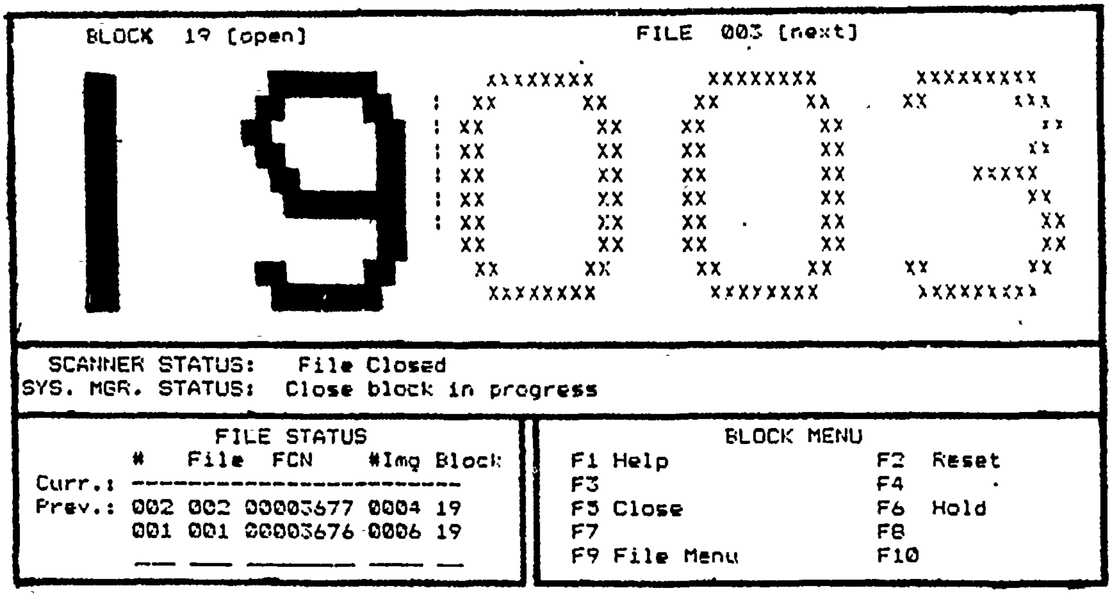
"File 002 Open"



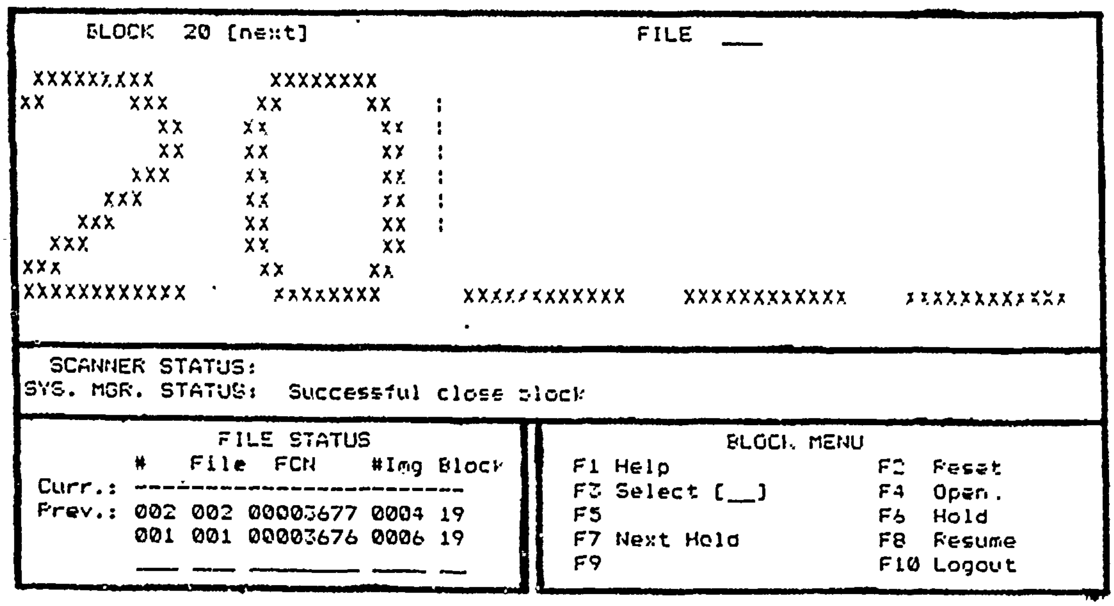
"File Closed"

Figure B-8

Operational Control Terminal; Block Closed



"Close Block in Progress"



"Successful Close Block"

Figure B-9

B.2.2 Low Speed Paper Scanners

ODISS has two paper scanners that operate without a paper transport. Each of these scanners utilizes a flat glass platen on which the original paper documents are placed. The primary differences in the two scanners are in the number of bits per pixel they scan and in the manner in which they perform image enhancement. In other words, one is a black and white scanner utilizing hardware[89] for enhancement, and the other scans in 256 levels of gray scale and uses software enhancement. Each scanner is described below.

<u>Black and White Scanner</u>

The black-and-white low-speed scanner is an IS-400 platen type scanner built by Ricoh. It is a desk top unit, that performs the function of converting images into digital data. This scanner has the capability to scan oversize documents up to 11" by 17". It is also especially useful for capturing fragile documents because the document is not moved or handled by the scanner's mechanisms. A novel feature of the scanner is its capability to scan at any resolution between 200 and 400 dpi. This capability is provided through the use of moving mirrors and lens which guide light reflected off the document image on to a 5000 element CCD array at different speeds and magnifications. Scanner software, limits the resolutions selectable to just three values, 200, 300 and 400 dpi.

Five mirrors in the low speed scanner system, divided into three mirror assemblies, each perform a specific task. The first assembly consists of one of these mirrors and two florescent bulbs. This assembly moves to pan a reflection of the entire document across the CCD in the subscan direction (down the page). The second mirror assembly consists of two mirrors mounted on a second traveler. This assembly follows the first assembly at one-half the first assembly's speed to maintain a uniform distance for the light path from the document to the CCD. The second mirror assembly has two mirrors, one of which is adjustable to focus the image on the CCD sensor.

The third mirror assembly has two mirrors in it, one of which is adjustable to focus the image of the document on the CCD sensor. The third mirror assembly also has the capability to reduce by 50% the size of the scanned image, but this feature is not implemented in the ODISS configuration. The lens system creates a reduced, focused image on the CCD array. The lens moves to a different position when the scanner is in 50% reduction mode. This movement ensures that the light reflected off the document is focused on the CCD. The drive system consists of three motors and their drive wires. The first motor has wires to move the first and second mirror assemblies, the second motor drives wires to move the third mirror assembly, and third motor (the lens motor) moves the lens via a focusing arm. A microprocessor with feedback from the mirror/lens assemblies provides digital control to all motors.

The electronics for the IS-400 includes motor control, the CCD sensor, analog and digital control circuits, a serial interface circuit, and the DC power supply. The components provide control for both the mechanical and data handling aspects of the scanner. The CCD sensor provides the high resolution scanning feature of the IS-400. Light from the florescent bulbs is reflected by the scanned document onto the CCD. An analog control board is used to calibrate the scanner for maximum efficiency by correcting the white level of the analog

[89] Hard-wired circuitry as opposed to software programs

scanner output. It also corrects inconsistencies in the intensity of light from the florescent lamps. The output of the analog control board is a 6-bit binary representation of the scanned document.

The signal from the analog control board is then digitally processed in one of three modes. Each mode is designed to give better results for a different type of scanned document. These three modes are character mode, photograph mode, and character/ photograph mode. Character mode puts the two signals through a modified transfer function (MTF). This has the effect of clarifying the picture by brightening the light areas and shading the dark area further. Photograph mode is used to increase clarification in documents containing half tones, and character/photograph mode is used to bring out both characters and half tone information in the same document.

There are only two connections made to the Ricoh scanner. These are for power and the RS-422 serial communication cable, which includes both data and control lines. The scanner is controlled from a remote keyboard at the scanner host.

Low Speed Scanner Operation

The low speed scanner position provides the following capabilities:

- Scan CMSR documents that have been marked for rescan by the Quality Control Workstations
- Scan CMSR documents that are physically unsuitable for scanning by the High Speed Scanner
- Scan Non-CMSR documents
- Receive images from the Image Enhancement Element (IEE)
- Store images within ODISS

The low speed scanner position consists of a Unisys PC/IT, display monitor, keyboard, and a Ricoh scanner. This position when operating and performing the capabilities listed above, is hereafter referred to as the Low Speed Capture Element (LCE).

Logging On

The log-on display is the first display presented to the operator after the LCE has been initialized by the System Manager. The operator enters a five digit employee identification number (EIN) in the highlighted field. The cursor automatically advances to the password field when the EIN field is complete. The operator enters an assigned password in the password field. While in this field, the cursor advances, but the characters are not displayed. The cursor then advances to the record type field. When all fields are complete, the operator should press the ENTER key. This sends log-on information to the System Manager.

The Mode Menu (Figure B-10) is presented in the same format, and is used throughout the log-on session. The upper right window is used to display miscellaneous information such as error messages. The lower right window is referred to as the status window. It is used to display the current scanner configuration as well as useful document scanning information.

Mode Menu

National Archives and Records Administration
Optical Digital Image Storage System
LOW SPEED SCANNER STATION

MODE MENU (press F1 for HELP)

F3 CMSR ENTRY MODE

F4 NON-CMSR ENTRY MODE

F5 CMSR RESCAN MODE

F6 SCANNER TEST MODE

F10 LOGOFF

Mode: none		Scanner Configuration
Block:		Density: 200 DPI
File:	FCN:	Size: 8.5 X 11
Page:	of	Mode: binary
		Threshold: 4
		Remove Texture: ON
Display: none		

Figure B-10

The middle window displays the current menu. The LCE is a menu driven system, in that all interaction is in the form of choosing available operations that are presented in a menu format. The operator chooses the desired operation by using the function keys on the left side of the keyboard. Table B-2 shows the options available from the Mode Menu.

Gray-Scale Scanner

The Xerox Inca-38 scanner is the capture engine for gray scale images. This scanner works on the same principles as the IS-400 scanner and, like the IS-400, is a platen scanner. The scanner can capture a document image with 8-bit gray level resolution at a linear resolution of either 200, 300 or 400 dpi. The scanner can scan images up to 11" by 14" in size. The scanner has one interface port on the rear of the device. The interface is a Xerox custom communication standard. Data presented at the port is unprocessed 8-bit image data, converted directly from the analog CCD output to digital data.

Mode Menu Options

KEY	DESCRIPTION
F1	Display Help commands.
F3	Go into CMSR Entry Mode and display the CMSR Entry Block Menu. This mode is used for scanning documents that have not been previously stored within ODISS.
F4	Go into Non-CMSR Entry Mode and display the Non CMSR Entry Block Menu. This mode is used for scanning documents that have not been previously stored within ODISS.
F5	Go into CMSR Rescan Mode and display the Rescan Block Menu. This mode is used for scanning document which have been previously stored within ODISS and marked for rescan by the Quality Control Workstations.
F6	Go into Scanner Test Mode and display the Test Scanner Menu. This mode provides a test environment for the LCE.
F10	Log-off the System Manager, return to the log-on display.

Table B-2

The companion Unisys PC/IT processes the 8-bit gray image based on a user interactive set of image enhancement algorithms. Once the image enhancement is completed, the image is binarized and packed for shipment into the core where it is treated as a rescanned image.

Gray-Scale Scanner/Terminal Operation

Once the operator has logged onto the image enhancement terminal and the system has verified the user profile that authorizes his or her actions, the operator is presented with a menu that indicates the specific function to be performed. This terminal is operated independently of the low speed scan control terminal as either support of document entry, analysis and resetting of procession parameters, or for demonstration. Operator choices in this menu are:

- Help
- Select file
- Transfer file to low speed station
- Designate region of interest
- Perform local image manipulations
- Perform disk file manipulations
- Perform scanning operations
- Perform image enhancement
- Reset default parameter values
- Redefine automatic processing
- Exit to DOS
- Exit to standard prompt

Image Enhancement Terminal Menus

The main menu tree for the image enhancement terminal is shown in Figure B-11. The menus are displayed on the color ASCII display attached to the image enhancement terminal and the information shown is a minimum that is provided to the operator.

Help Screens

Help screens are available from several menus in the system. These screens describe how to use the menu tree structure, what some of the image enhancement techniques do and how they are used, and how to operate certain functions such as erasure or selection of region of interest. Pressing the indicated help key causes a help screen to appear. The operator may step through the screen, one line at a time, using the UP ARROW and DOWN ARROW keys; jump a page at a time using the PG UP or PG DN key; or skip to the beginning of the file with the HOME key or to the end of the file with the END key.

The operator may specify a new image name, length and width in inches, and density (200, 300, 400 dpi); or the operator may specify an existing image, in which case the parameters stored with that image are used.

Main Options Menu

ODISS IMAGE ENHANCEMENT - MAIN OPTIONS
Alt/F1 - Help
Alt/F2 - Select File, Set Parameters
F1 - Transfer File to Low Speed Station
F2 - Designate Region of Interest
F3 - Local Image Operations
F4 - Disk File Operations
F5 - Scanning Operations
F6 - Image Enhancement
F7 - Reset Default Parameters
F8 - Redefine Automatic Processing
F9 - Exit to DOS
F10 - Exit

Figure B-11

Transfer File to Low Speed Station

The image enhancement terminal packs the one bit per pixel images resulting from the enhancement process, including thresholding or halftoning, and sends the image to the high resolution low speed scan control terminal for display. Alternately, the processed image may be temporarily stored on local storage for later transmittal to the low speed scan control terminal. Operator choices in this menu are:

- Select Binary Image

  If the image has already been bit-packed, the binary image is selected.

- Binarize and Pack Image

  Packs black & white image so each pixel is represented by one bit (8 pixels per byte). This is the format required for display on the low speed station.

- Transfer Image to Low Speed Station

  To transfer a binary (bit-packed) image to the low speed station, the Low Speed Station operator must select "RECEIVE IMAGE A/B from Image Enhancement". The image enhancement terminal operator may then select "Transfer Image to Low Speed Station". A window then appears at the bottom of the screen with the message "Transferring Header Block", followed quickly by "Transferring Image Block # 1", "Transferring Image Block # 2", etc. The total number of blocks transferred depends on the size and density of the scanned image.

Designate Region of Interest

The operator must designate whether image enhancement algorithms are to be applied toward the entire scanned image, a small experimental image, or a "region of interest" selected from the entire image. If region of interest is chosen, a subsample of the image is created and displayed. A rectangle representing a screen-sized area of this region appears on the screen with the subsampled image. This rectangle may be positioned anywhere on the image, using the arrow keys, until the desired region of interest is designated. At this point the selected region of interest is displayed at full resolution. The purpose of this step is to provide the operator with the option of specifying a region of interest for subsequent enhancement while selecting ranges of contrast stretch, for example, or by varying other enhancement processing constants.

The operator soon becomes adept in analysis and selection of the processing operations and parameters that are useful for correcting various document deficiencies. The skilled operator may use this step to select the specific operations to be performed over the entire document, or may bypass this step and proceed immediately to process the entire document with standard processing sequences. The image enhancement processes available to the operator are described fully below.

Local Image Manipulations

Local images are those stored in high-speed memory in the image enhancement terminal processor. These images are usually regions of interest measuring 512 x 480 pixels. The

images are generally those that are being used by a demonstration operator for study of image enhancement processes and the effect of changes in parameter values or the order of image enhancement processing. Local image manipulations include:

- ☼ Rename local images
- ☼ Copy local images to disk
- ☼ Save gray screen as local image
- ☼ Delete local image
- ☼ Display local image on gray screen
- ☼ Display local space available

All images to be accessible to these menu controlled processes are restricted to the I:\IMG directory in high-speed memory. Image names are limited to eight characters, without extension.

Available disk file operations are described below.

- ☼ List all images recorded on system: lists those files for which a header file exists, regardless of disk partitions. Header files are created and maintained by the system and represent complete scanned images.

- ☼ List images currently on disk- #xx: lists all ".img" files on the specified drive. This includes scanned images as well as local images (areas of interest) saved on disk.

- ☼ Select disk drive: allows the operator to select from the following drives (partitions): "c:" "d:" "e:" "f:" "g:" "h:".

- ☼ Display local space available: displays the number of bytes available in RAM extended memory.

- ☼ Display disk space available: displays the number of bytes available on the specified drive (partition).

- ☼ Copy local image to disk: copy local image (area of interest) to specified drive (partition).

- ☼ Copy disk image to local: copy disk image (area of interest) to RAM extended memory.

- ☼ Copy disk image to disk: copy disk image to new name, drive, or both.

- ☼ Delete disk file: allows deletion of disk files for which headers exist.

- ☼ Display disk file: allows operator to display subsampled image, then select a region of interest for display and further processing.

Scanning Operations

Scanning operations include:

- Setting the size of the image to be scanned and the scan density.

- Starting the scan process.

Upon completion of the scanning process, the entire image is subsampled to allow the entire document to be displayed on the attached gray scale monitor in 512 by 512 format. The full scanned image is stored on disk in the designated file.

Save Current Series

The operator is allowed to save the current sequence of specific enhancement processes. This saves those processes executed after the most recent invocation of F6.

Perform Default Processing Series

Selection of one of the two predetermined sequences of image enhancement operations causes the designated total image, experimental image, or area of interest to be processed by the predetermined sequence using standard default parameters. The operator selects the desired sequence and has no further input until the process has been completed and the results are displayed on the gray scale display. The specific nature of the sequence is determined through experimentation with many of the damaged documents in the database.

B.2.3 Multiformat Microform Scanner

The multiformat microform scanner, built by Photomatrix Corporation, is designed to scan microfiche, aperture cards and 16mm and 35 mm microfilm. The scanner consists of two sets of components, a scanning component and an electronic image processing component.

The scan component of the multiformat scanner is an adaptation of the unit found in standard Photomatrix microfiche scanner. The transport mechanism is altered to accept the variety of microform media. This modification allows manual positioning of a single document image in the scan window of a CCD array. Each type of microform media requires a mechanical adapter specially designed to position and hold the media properly. For reel film media, input and takeup reels are provided.

The adapters provide an edge to aid in the orthogonal alignment of the media with the scan window. The operator first positions the document image in the vicinity of the scan window of the CCD elements. Fine positioning of the image with respect to the window is accomplished by the operator through use a host computer keyboard to control the x-y[90] movement of the media transport mechanism. The x-y movement can be as fine as a single CCD array element. This same movement mechanism is also used to scan the entire image automatically after the operator has made his fine positioning adjustments.

[90] Horizontal and vertical

The scanning operation moves the x-y mechanism past an array of three CCD arrays, each with 3860 elements. This is sufficient for a 400 DPI resolution on an 8.5" x 11" document reduced by 48X. The optics of the scanning mechanism split the light transmitted through the film to accommodate the physical positioning of the CCD arrays, allowing each array to take a separate optical position along a single scan line. The digital image data signals from the three CCD arrays are sent to the electronic circuits where they are written to a buffer. Since the data from the three physical sensors may be slightly skewed, the compensating horizonal and vertical offsets are located here.

The output scan lines are stored, five lines deep, in a buffer that is used for image processing. The processing hardware used in the microform scanner is similar to that used in the high speed scan and the same processing capabilities are available to the operator.

The microform scanner communicates with external equipment over three different interface lines. Two of these lines are for control information and are supported at DB-25 female connectors on the rear panel of the microform scanners. One of these ports is labeled "terminal" and can be used for direct control from a dumb terminal interfaced to it. The second port is labeled "host" and can communicate with an intelligent controller. This pair of ports conforms to the RS-232 interface standard. The third communication port is also supported on a DB-25 connector, and is used to transfer image and status information to an image interface. The signals passed on this line conform to the RS-422 electrical interface standard. This port is labeled "image data" and is also on the rear panel of the microform scanner.

B.3 Image Enhancement and Quality Analysis

<u>High Speed Image Enhancement Control Operation</u>

There are three ways to control image enhancement automatically on the high speed scanner: (1) fixed thresholding; (2) dynamic tracking; and (3) image processing thresholding.[91] Image processing thresholding is discussed here since it is easiest for the operator to control.

The high speed scanner provides some image enhancement through the LIGHT and DARK buttons found on the front panel. These buttons, used in conjunction with the NORMAL button, allow the operator to adjust the darkness of the scanned image. The processed threshold is activated by pressing the DIAG SELECT button until the PROC THRESHOLD message appears, then press the DIAG EXEC button until PROC THRESHOLD again appears. Now the normal button selects VT TOP CCD = nn or VT BOT CCD = nn. With the top or bottom CCD selected, the LIGHT and DARK buttons make the corresponding image lighter or darker. TOP and BOTTOM refer to the camera positions. The top camera reads the backside of the documents when fed face down. Successive LIGHT or DARK button activation makes the images lighter or darker until the numerical limits are reached.

<u>Low Speed Image Enhancement</u>

In the later stages of the ODISS test, Image Processing Technologies, Inc., loaned NARA a new image processing device that they installed in the Ricoh scanner. The "Scan Optimizer" is a board-level device that is installed on the scanner's mother board. The enhancement

[91] See Appendix A for a complete discussion of thresholding techniques.

activity is regulated by a hand-held controller that is interfaced with the Low Speed Workstation. All the operations of the Scan Optimizer work serially with the normal functions of the Low Speed Workstation. For instance, changing parameters on the workstation such as moving from 200 to 400 dpi would have no effect on the Scan Optimizer and vice-versa.

The hand-held controller contains a keypad and an LCD display. It is connected to the system unit by its cable. The keypad on the hand-held controller is used to change the Optimizer parameters. These parameters vary the image produced by the scanner, changing line thickness and sensitivity to fine details. The keypad also provides access to some diagnostic and initialization features. The two lights on the controller indicate the status of the unit when properly initialized.

The hand-held controller has a two-line display to provide information about the optimizer. The top line is the command line. It is used to display messages from the Optimizer and to show new parameter values as they are being entered. The bottom line is the status line which displays the current parameters. This line will be blank if the unit is not enabled. In this case, the normal Ricoh scanner parameters represent the only enhancement actions employed to clean up the document.

The individual parameters available on the optimizer are the following:

- <u>ENTER</u>:
  The ENTER key completes an operation, such as changing a parameter. After the new value has been selected, the ENTER key is pressed to store the data.

- <u>CANCEL</u>:
  The CANCEL key aborts a parameter change, leaving the parameter at its previous value. This is useful if a parameter key is pressed accidentally.

- <u>PRESET</u>:
  The PRESET keys set each of the OPTIMIZER's parameters to a given value. There are ten presets which are accessed by pressing a number key (0 - 9) when the mode light is lit. The parameters stored in a preset can be changed using the SET key. Different presets can be defined for each type of document to be processed so the proper parameter values can be set by pressing only one key.

- <u>SET</u>:
  The SET key stores the current parameter values to one of the ten available presets, overwriting the previously stored values.

- <u>LINE</u>:
  The LINE key changes between outline mode and solid mode. In outline mode, only the edges of regions are shown, while in solid mode, these regions are filled.

- <u>THK</u>:
  The THK key controls the thickness of the spaces between the border lines created by the outline mode. The parameter values range from 1 - 31 with 31 representing the thickest space. When a thin value is used in conjunction with the outline mode, the resultant figure is well defined with little or no space between the borders.

* BRI:
  The BRI key alters the brightness of the filled areas of the image as opposed to the overall brightness of the entire image. The value can be set between 1 and 4, 1 producing the lightest output.

* INV:
  The INV key changes between inverted and normal mode. In inverted mode, black areas on the document are shown as white and white areas are shown as black. In normal mode, the black and white areas remain unchanged. If the outline option is selected, this key can be used to place the edges inside a black region (normal mode) or inside a white region (inverted mode).

* FLTR:
  The FLTR key changes the amount of filtering performed on the document image. The value can be set at 0, 1 or 2, with 0 indicating no filtering. Using additional filtering will help eliminate stray dots within the background of the image. It may, however, cause rounding of the corners of objects and cause very thin lines to vanish.

* WSIZE:
  The WSIZE key changes the working window size between 5 x 5 and 7 x 7. This window indicates the area of comparison of contiguous pixels within the image for the purpose of enhancement control. A 7 x 7 window size produces thicker lines but eliminates more stray dots than the 5 x 5 window size.

* SENS:
  The SENS key changes the sensitivity of the overall document image. This controls the level of brightness for the entire image instead of just the character information as in the brightness control. The value can be set between 1 and 31, with 31 generating the lightest results. Lower sensitivity settings are used with poor contrast documents. Higher settings are appropriate for documents with colored or otherwise dark backgrounds.

Gray-Scale Image Enhancement Operation

The operator may choose to perform image enhancement on the entire document, or may first evaluate the effect of the various enhancement processes on a small portion of the document (region of interest measuring 512 by 512 pixels). The advantage of the latter approach is that much more rapid response is provided to each operation and the operator views the results almost immediately. After the processing has been performed on the small region of interest, the operator may apply the same sequence of processes to the entire document.

Specific Gray-Scale Image Enhancement Processes

* Calibrate. The calibrate function provides for the radiometric correction to the image to compensate for nonuniform illumination of the image during scanning. Two additional images must be available to the operator in order to perform this function: a flat field image and a dark current image. Each of these images must have been scanned in the same manner as the active image. The number of dots per inch must be identical. The images must be recent enough such that the illumination source has not degraded since the scan was made.

- Linear Contrast Stretch. The selection of the linear contrast stretch function provides the operator with the option to specify the lower and upper percentage saturation parameters. The lower cutoff value is the percentage of the darkest pixels which are to be set to black, the upper cutoff is the percentage of the lightest pixels which are to become white. All pixel intensities lying between these two limits are "stretched" to lie from black to white (0 to 255). For example: Specifying a stretch with a lower cutoff of 10% and an upper cutoff of 1% causes the darkest 10% of the pixels to become black (value 0) and the lightest 1% of the pixels to become white (value 255). Of the remaining pixels, the dark become darker and the light become lighter until the image is represented by values all the way from 0 to 255. The resultant image is displayed on the gray scale display at completion of the processing.

- Convolution Filter. The convolution filter function allows the application of a convolution filter to the image. The operator may select the size of the filter to be used (size of the matrix where a 7 x 7 matrix is the normal maximum, but the system allows selection of up to a 9 x 9 matrix). The operator then has the option of selecting the scaling parameters to be used, and to select the filter weights. It is expected that a set of default values will be generated through experience, and that the operator will then select and input the appropriate set. The resultant image is displayed on the gray scale display at completion of the process.

- Unsharp Masking Filter. Selection of the unsharp masking filter function causes the system to apply a high pass filter with an unsharp masking constant to the image. The program automatically computes values for the scaling constant during the setup and displays this value to the operator. The operator may accept this value, or may override the value and input another scaling constant. The operator may also specify the value of the unsharp masking constant, or accept the default value. The image resulting from the process is displayed on the gray scale display upon completion of the process.

- Edge Detection. Selection of the edge detection process next requires that the operator select either the Sobel detector, the Roberts detector, or the LaPlacian detector. At the completion of processing, the results of the process are displayed on the gray scale display.

- Adaptive Filter. Selection of the adaptive filter procedure allows the operator to input values for the size of the neighborhood of gray values that to be averaged, but the scaling constants cannot be changed. At the completion of the process, the image is displayed on the gray scale display.

- Threshold. Selection of the threshold function allows the operator to apply dynamic thresholding to the image, or to apply a constant threshold to the entire image. At this time, the operator may select the value of the constant that is to be used for the value of the threshold. At completion of this process, the completed image is displayed on the gray scale display. The image in its entirety may also be sent to the low speed scan control terminal for viewing in monochromatic high resolution form.

- Halftone. The selection of halftoning causes the image to be dynamically processed with a halftone algorithm based on local area statistics. The operator must select

whether a 2 x 2, 3 x 3, or 4 x 4 halftone pattern is to be used. No parametric values may be changed by the operator. Results of the process are displayed on the gray scale display. The image in its entirety may also be sent to the low speed scan control terminal for viewing in monochromatic high resolution form.

Image Erasure on Gray-Scale Workstation

The selective editing of documents takes place after completion of all other processing on the image. This capability is provided for the final removal of flaws in the document which are not removable by other means. Also, the erasure process takes place on images that have already been thresholded and, except for bit-packing, are ready for transmittal to the low speed scan control terminal. The first operation is to determine from the full sized document image the region of interest for subsequent mark removal. This region of interest is a square area 512 by 512 pixels for full resolution display on the gray scale monitor. Alternatively, the operator may elect to step through the entire document, in 512 by 512 steps, erasing in each area as he or she proceeds.

Two forms of selective editing are provided. First, a small rectangular area of interest may be drawn and moved with cursor control to surround an area in which all mark removal is desired. When the operator has placed and sized the rectangle, he or she then causes erasure by use of the function key. Alternatively, the operator may select an icon with an "eraser" tip. The eraser may be turned off and on under operator control. When on, any areas under the tip are erased and the bit indication changed from black to white. When a given region of interest has been completely edited, the operator may select another 512 by 512 region for additional editing prior to completion of his or her work. The final document is not available for further gray scale enhancement until after the erasure process is completed.

B.4 Production Throughput Capabilities

In order to size the original system design properly, Unisys required an analysis of the sample documents, file structure and the throughput required of the conversion subsystem. Based on the given information, Unisys devised a system concept that, in their opinion, would meet the required goal of 15,000 images converted per day. This estimate was couched in several caveats, however. The files and documents had to be exactly as described and all other circumstances needed to be ideal.

If these criteria were met, Unisys thought that they could process the required 15,000 images per day in order to convert the 1.2 million images of the sample within four months. The TDC 4200 high speed scanner's rated speed was 35 pages per minute. The allotted time for conversion of 1,000,000 pages was 560 hours. The equivalent of 1786 pages had to be scanned every hour of the 7 hour work shift. This translated to 30 pages per minute. The TDC 2000's rated speed exceeds this figure, so the scan rate was theoretically feasible. External image manipulations such as image data compression and thresholding would take place within the real time of the scan thereby exacting no extra time of any significance.

The scanned images would be temporarily stored on the image capture magnetic storage buffer where they would remain throughout the indexing, quality control and rescan process. Any workstation or combination of workstations within the system could then be utilized for indexing. Once the file was indexed, block level batch processing of the remaining steps of quality control and rescan was possible. A typical deployment scenario would have one high speed scanner scanning at the rate of 30 pages per minute (two files at the estimated size).

One index station could easily handle key entry for two files in one minute. Two stations designated for quality control could review the index data and image presence and quality for one file of 15 pages or 18 images (83% two-sided pages x 15 pages = 18 images). This appeared to provide the most efficient assignment of activities for the conversion input function.

B.5 Indexing

Creating the CMSR index is the second major step in the input process. A file control number is assigned when the file is opened at the scanner, but the subject terms in the index record are generated at the indexing work station. The index station is not usable for non-CMSR records; it is used only for CMSR files.

B.5.1 Subject Terms For CMSR Files

The subject terms attached to each file at the index station allow searches and retrievals based on the researchers' knowledge of the names of the soldiers, of the regiments and companies in which they served, and of their ranks.

The first, middle, and last names of the soldier are three separate alphabetic fields. Numeric code table values are used for the other fields. There are 204 numbers for the regiments of the Tennessee confederates, and for each regiment there are tables that allow entries for up to three companies within the regiment. A table of thirteen code values covers the starting and final military ranks (ranks in and out) of the soldiers.

There is also an alphabetic Remarks area for transcribing cross references to other relevant files that sometimes appear on the files' jackets. Unlike the other fields, no searches or retrievals can be performed on the Remarks.

B.5.2 Station Workflow

Files are available for indexing as soon as they have been completed at the high speed paper scanner. They arrive at the index workstations in a somewhat random way that approximates first in/first out order.

The paper records are not needed at indexing since the first image appearing on the screen when a file arrives at the index station is the jacket, and this image contains all the information needed to build the file's index record. Occasionally some piece of information may be missing from the jacket, and in these rare instances the operator can page into the file to look for the information on the other images.

The images appear on the left side of the terminal's screen, while prompts for entering the index terms appear on the lower right side. The upper right side of the screen is an area for instructions. By using a function key, the operator can get lists of the numeric code table values for the regiment, company, and rank fields in this upper right side section.

Reading the information from the images on the left side of the screen, the index operator fills in the index fields in the lower right section in the following order. last name, first name, middle name, rank in, rank out, regiment, companies, and remarks. No field can be left empty; so when there is no information for a numeric field, a zero is entered, and when there

is no cross-reference notation on the jacket, the word "None" is entered in the remarks field. If the operator notices a mistake in a field, it can be corrected by rekeying the entry.

After the index for the file is completed, the operator presses a function key that enters the file's index record into the database and removes the file from the workstation. At the same time, this action calls the next file to the workstation with the new file's jacket appearing on the screen.

B.5.3 Hardware Configuration

The index workstation consists of the same essential hardware used for the quality control, demonstration, staff retrieval, and public retrieval modes, and this description applies to each of these other workstations. The modularity of ODISS allows any number of the eight workstations in the system to be allotted in any combination to the functions of indexing, quality control, and staff or public retrieval. Refer to Figure B-12, for a diagram of the general workstation architecture. Two workstations were assigned to indexing in the initial plan for CMSR processing, but any other workstation can be dedicated to indexing to deal with backlogs.

The stations consist of Sperry PC/ITs with Discorp image processing boards and video display monitors. The monitors have a 19-inch diagonal. They display images in black and white. For indexing and the other major workstation modes, the display screen is split between an image display area on the left side and an alphanumeric display area on the right. In this standard display mode, images are shown at 150 dpi, but there is a capability to display at the original scan resolution of 200 dpi, and a 2x zoom is also available. Documents up to 8.5 x 11 inches can be displayed at full size in the standard screen display mode.

Three workstation servers in the core cabinet are Heurikon HK68/M10 single-board computers. The Heurikons are based on the Motorola 68010 microprocessor that has a clock speed of 10 Mhz and a multitasking capability to permit the computers to perform two or more tasks at the same time. Two of the Heurikon processors serve three workstations, while the third processor serves two workstations.

Each Heurikon server is supported by a 170 MB Maxtor XT-3170 disk for temporary storage of image data and files. Each disk has 170 MB capacity unformatted and 132.5 MB capacity formatted. Communications between the workstations and the Heurikon servers in the core cabinet operating under the central System Manager are handled on hard wires. An RS-232 line is used for non-image data, and an RS-422 signal cable transmits image data.

B.5.4 Software Capabilities

Unisys's software written in the C language controls the workstation activities, such as the menus for indexing and the display of code tables. Unisys's software also implements the modularity that permits the user to access the different functional modes, such as indexing or quality control, of the workstations. The software provides the links between the different components of the system and coordinates the workstations' activities with the other software modules such as the printer module for generating hard copies and the system manager module to update or query the CMSR database.

Workstation Subsystem

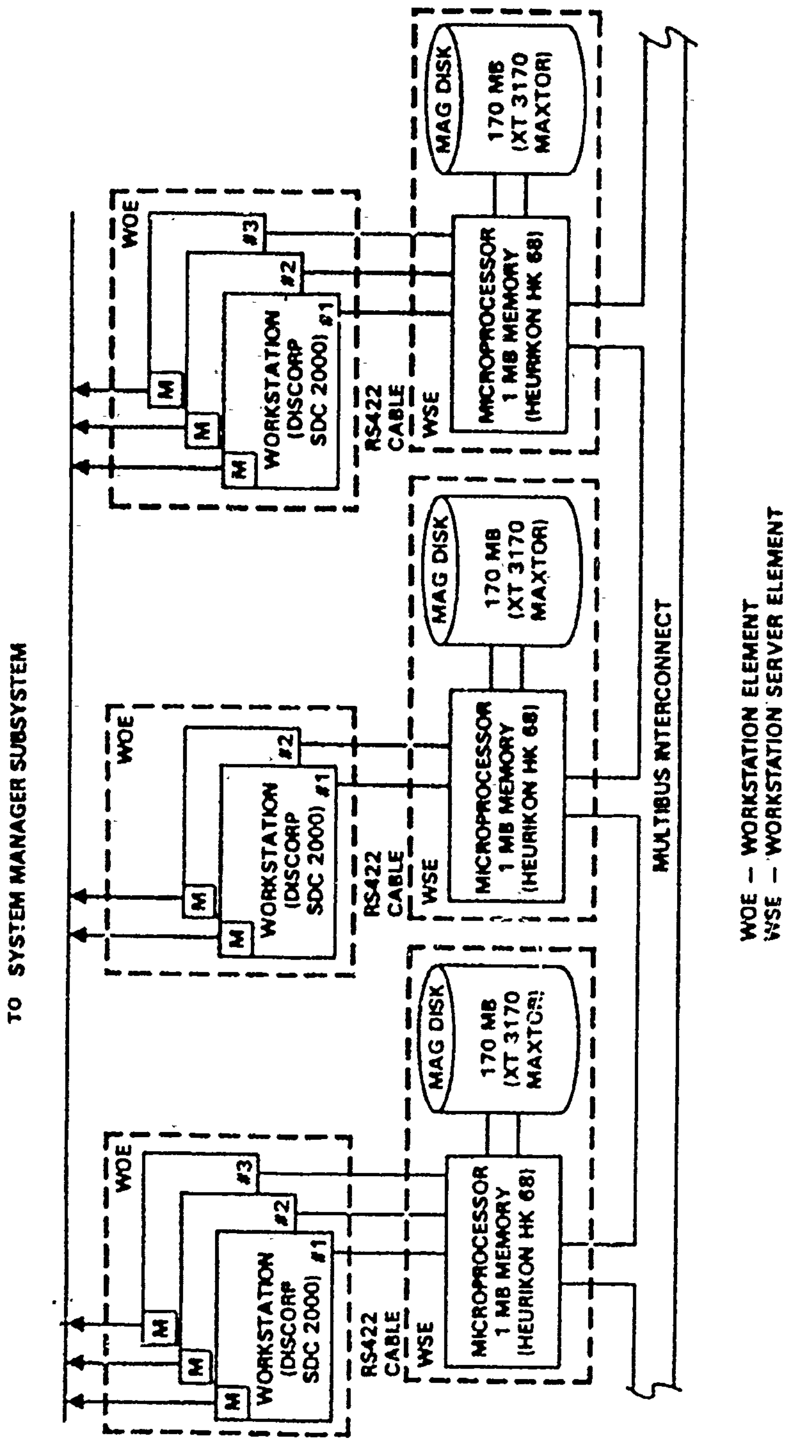


Figure B-12

The actual index records created during indexing for the CMSR files are stored in, and subsequently accessed from the CMSR database established under the UNIFY database system as part of the system manager software module. The UNIFY relational database management system operates within a UNIX environment.

ODISS utilizes three operating systems. The workstations run under MS-DOS while the Heurikon processors in the core cabinet run under the VRTX operating system and ODISS as a whole runs under UNIX. The links between the different components and the coordination between their different operating systems are provided by Unisys software. Most of the software created by Unisys was written in C, but some was written in assembly language in an effort to maximize performance.

B.6 Quality Control

Quality control is the third major input step. Only CMSR files could go through the quality control stage. Non-CMSR files could not be sent through the quality control process.

B.6.1 Purposes

There are two purposes for quality control. The first is to catch any mistakes made at indexing, such as misspellings of names or entering the wrong numeric code value for a regiment, company, or rank. The quality control station gives the operator the capability to change the index entry in any field so that mistakes can be corrected.

The second major purpose is to review the files for image quality. The station gives the operator the ability to mark poor images for rescanning and image enhancement at the low speed scanner/rescan station. This review to improve image quality has been made of all the documents in all the CMSR files converted to digital format.

B.6.2 Station Workflow

The CMSR records arrive at the quality control station in blocks that typically consist of 40 to 60 CMSR files. These are the same blocks that were created at the high speed scanner at the beginning of the input process. The blocks normally consist of all the files that were placed in one Hollinger box during document preparation, and their equivalent digital image files created during initial scanning.

At quality control the operators work with both the paper files and the digital image files. After signing onto the system, the operator sees a list of available blocks on the terminal's screen, selects one block for work, and gets the corresponding box of paper files. Once a block is selected, the files come to the screen in consecutive order, which usually is from last to first (i.e., nos. 60, 59, 58, ... 3, 2, 1). The images appear on the left side of the screen and the index fields with their completed entries appear on the lower right side of the screen.

Files come to the screen after the quality control operators press the appropriate function key. In each file, the images appear on the screen in page order from first to last. Each succeeding image comes to the screen when the operator uses the PAGE UP key, and, if necessary, the preceding image can be recalled with the PAGE DOWN key.

The jacket as the first item in every file always appears on the screen first. The operator selects the matching paper file and compares the index information on the jacket with the

index entries on the lower right side of the screen. If there is a need to look at any code table, such as the company table which can be unique to each regiment, the table can be brought to the upper right side of the screen by pressing a function key. When an indexing mistake is seen, the operator corrects the error.

As the operators are checking the accuracy of the indexing, they also evaluate the legibility of the jacket's image as the first item in the file. After finishing the index check the operators proceed through the file in consecutive page order comparing each image with its corresponding paper page.

If an image is illegible or otherwise is of poor quality, the operator uses a function key to mark the image electronically for rescan. At the same time the corresponding paper page is placed in a bright colored folder, and then the colored folder is placed back into the file's folder at the proper location. If there is no image for a page because the page was missed inadvertently during scanning, the paper document is put into a colored folder and returned to the file's folder. The proper location in the digital file is marked electronically with a Missing Page indicator, which will tell the rescan operator where to insert an image of the missed page.

When the review of the file is completed, the operator presses a function key that removes the file from the screen, builds a table of poor images for rescan and missing images for insertion, and retrieves the next file in the block to the screen. When the last file in a block is completed, the operator can return to the initial quality control menu to select another block or can log off.

B.6.3 Hardware Configuration And Software Capabilities

The quality control workstation has the same hardware as the indexing station or any of the other functions that are performed on the common ODISS workstation: video display monitors, Discorp image processing boards, a Sperry IT CPU, and RS-232 and RS-422 cable links to Heurikon processors in the core cabinet.

The software for quality control is essentially the same as that described for the indexing workstation. Software written by Unisys permits accessing the UNIFY CMSR database for blocks and individual files. The quality control menus and function keys operate under Unisys programs, and the blocks of files are accessed through the system manager software module. Two stations are assigned to quality control in the initial work plan, but the modularity of ODISS allows the assignment of additional workstations when backlogs occur at quality control.

B.7 Digital Storage

Index and image data captured during routine ODISS operations requires in-process and long-term storage. The ODISS system includes two categories of data storage: magnetic disks for temporary and permanent index and image data storage and optical disks for permanent image data. The document images captured at the scanners, along with the Ascii information keyed in at the index stations, are stored on magnetic disks in the capture storage element. Magnetic disks are useful since they allow correction of incorrect index data or poor quality images. For searching speed, the finalized index data is permanently stored on magnetic disks. This data was also copied daily onto streamer tapes for backup security. Optical disks store the actual CMSR image data, using write-once Sony optical disks. Each optical disk

holds up to forty thousand compressed images, any one of which is easily retrieved through simple workstation menu commands.

ODISS is designed so that a station production operator or a system user need not be concerned with the storage media used. The user has to only request the next file, or search of the database, and the system automatically will perform whatever task is needed. Image quality, or access to images is not limited in any way based on storage media. The following paragraphs describe the storage technologies in greater detail.

B.7.1 In-Process Image and Index Data

This data storage is identified as in-process, since it is a major component of the index and image data storage processes. The capture server element is a single-board computer (HK68/10). This computer accepts image data from three different sources. The element provides temporary data storage for image data before it is recorded onto optical media by the Archive Subsystem. Data from the film scanner and low speed scanner controller come over RS-422 interfaces. See Figure B-13 for a block diagram of the capture subsystem.

Data from the high speed capture element are placed into capture server disk storage by the on-board DMAC, using memory to map to the location of the image data in the multibus window. The three drives combined provide 840 (unformatted) megabytes of temporary storage for image data captured by the ODISS scanners. This storage provides a staging area for the data before they are irreversibly recorded onto the write-once optical media. This memory can hold over 700 average size files. The files are held in this memory until every file in an entire block has passed the quality control function and are written to optical media. Maxtor hard disk drives model XT-3280 (280 MB unformatted capacity) are used in the capture storage function. The recording media are 5.25-inch disks coated with magnetic material. Each disk has a movable read/write head that accesses and records information. The heads are positioned by a voice coil activator, which provides operation under high-density (1070 tpi) track spacing. The disk drive enclosure conforms to the standard for Winchester disk drives, and the power connection and DC voltages also conform to industry standards.

The data on the disk drives of the capture server element can also be sent as image data to the other ODISS subsystems. These data are usually the image data source for the operations performed by the index and quality control workstations. The index information from the System Manager is passed over the multibus to the capture storage element. This index information is interpreted by this element and, if the image data are on the magnetic disks, retrieved. This index information is then sent over the multibus to one of the workstation server elements. Images can also be printed from these data storage file locations.

B.7.2 Optical Disk Archival Storage

The ODISS optical storage devices are capable of storing the amount of information that is needed to handle a volume document imaging system. Without the storage density available on optical disks, the storage needs of NARA would make ODISS implementation unrealistic. The optical storage systems used in the ODISS configuration are based on the 12" write-once media adopted as standard by the Sony Corporation. There are three devices in the ODISS system that provide functionality for using these storage media. The first is the WDD-3000 writable disk drive, which writes digital information to the disk media. The second is the

Capture Subsystem

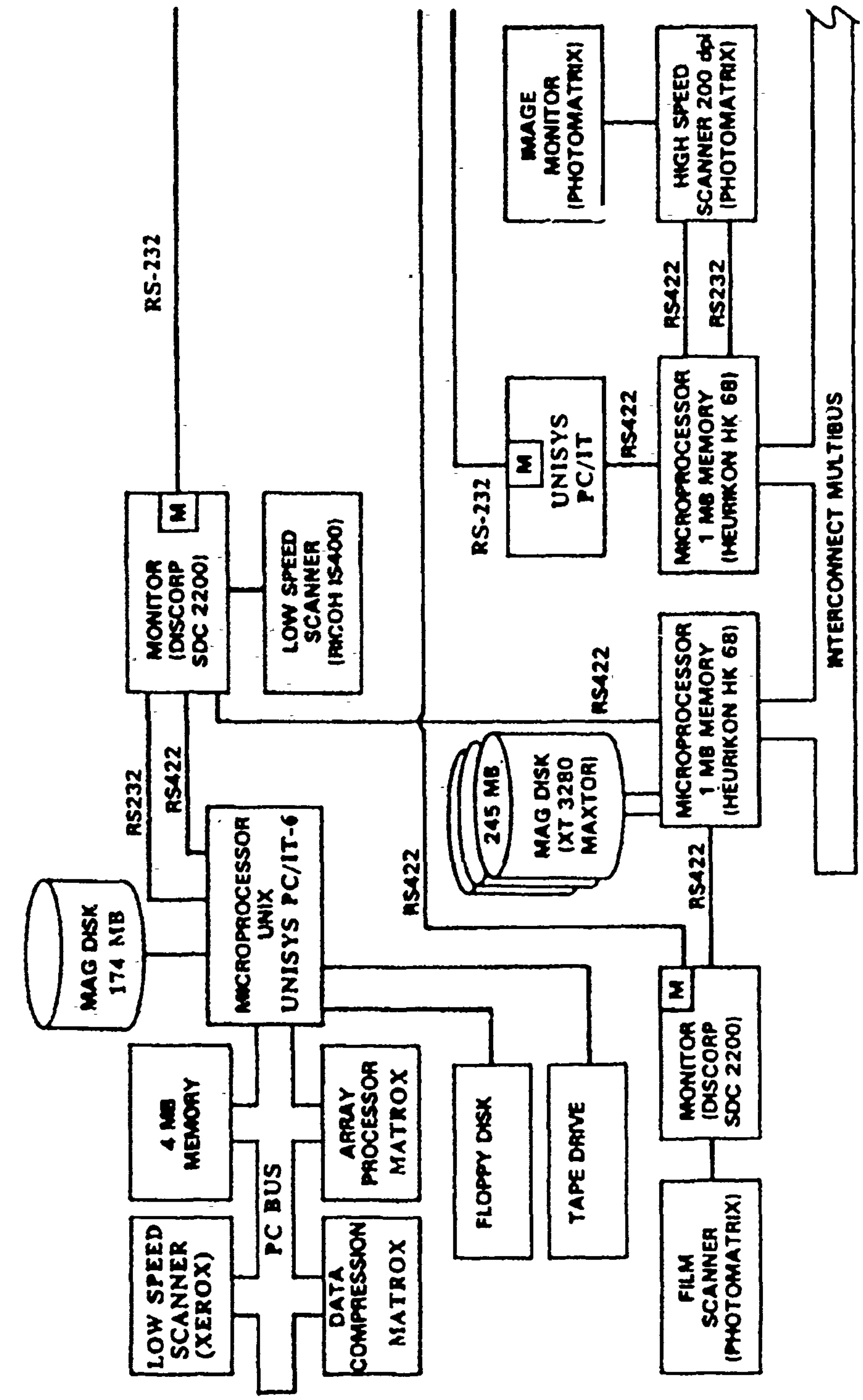


Figure B-13

WDC-2000 writable disk controller that provides direct interface for control of a writable disk drive from a host computer. The third is the WDA-3000-10 writable disk autochanger (jukebox), which provides a mechanical writable disk storage and retrieval function. Because the operation performed by this device is logically similar as the jukebox record players, this device is marketed under the name of the Sony jukebox. See Figure B-14 for a block diagram of the Archive Subsystem.

WDC-2000-10 Writable Disk Controller

The Sony WDC-2000 series controllers are writable disk controllers used to control the actions of the Sony WDD-3000 optical disk drives. Two models of the controller are in ODISS, the differences being in their interface to the computer. One controller is located in the jukebox, and the other is physically located within the core hardware rack enclosure. The internal operation of both controllers works to arbitrate instructions and communications to and from the host, a Heurikon HK68. Each controller can control up to 8 writable disk drives. The drives are connected to the controller via a fifty line cable connected in a daisy chain manner.

WDD-3000 Writable Disk Drive

The Sony WDD-3000 is the disk drive for writing and reading the Sony optical disks. There are four WDD 3000 drives within the ODISS system. Two are within the Sony jukebox, and the other two are in the System Manager hardware cabinet. These writable disk drives contain lasers to write to and read the optical disks. The WDD-3000 has the ability to record information on the write-once optical disks used in them. Information is recorded by heating the area designated for a particular bit to be set. This causes a phase change in the substrate layer of the disk. This new phase has a different refraction index from the original surface that allows its detection by interfering with the reflection of a read laser incident upon that section of the disk.

WDA-3000-10 Writable Disk Autochanger

The Sony WDA-3000-10 is a writable disk autochanger (jukebox) with an SCSI interface. This autochanger is used in conjunction with a Sony writable disk controller and writable disk drives to create an optical disk mass storage system capable of recording and storing up to 164 GB of user information on 50 disks. The autochanger controls the mechanical transportation of Sony optical disks to and from racks used for storage and drives used for recording and reading data on the disks. It also accommodates the controller and drives within its structure.

WDM-3DAO Writable Disk Media

The optical media used in ODISS is a 12 inch disk that is recorded and read at a constant angular velocity (CAV) of 720 rpm. A single disk can hold up to 1.091 GB of data. The disk substrate is sealed to the coating for long media life. The media is write once, read-many times, and the recording process causes an irreversible change to take place that creates permanently recorded data.

Archive Subsystem

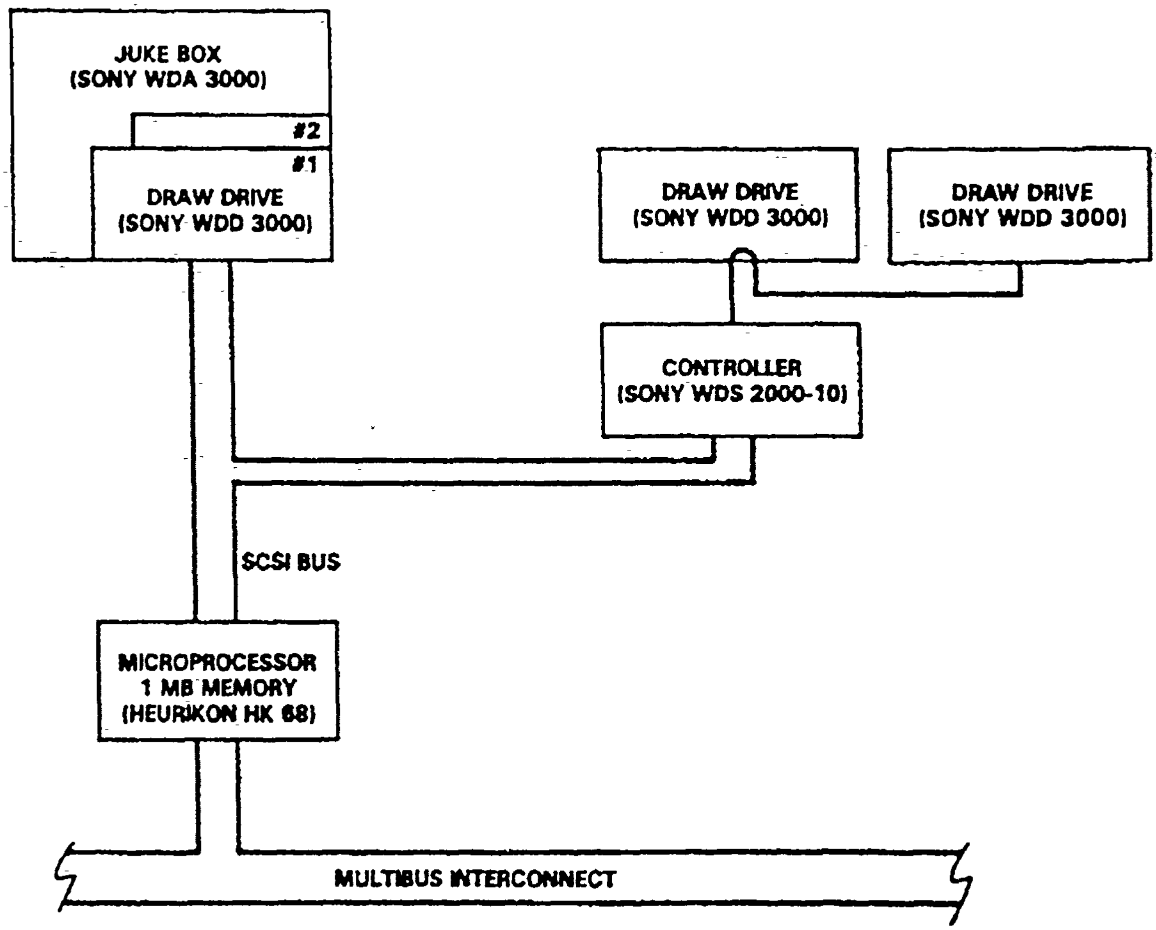


Figure B-14

B.7.3 Archives Subsystem

The Archive Subsystem performs the archive function for ODISS. It also records image data for the entire ODISS system. After a complete block of files is accepted by the quality control function, it is presented to the System Manager for archiving. The System Manager initiates the archive function by sending a command to the Initiation and Monitoring Subsystem. The image data are then sent to the Archive Subsystem from the capture storage element and are recorded onto optical media by one of the drives in the Sony jukebox or one of the external drives.

When a file is archived, the disk volume ID, disk side, and starting sector number are part of the identifying parameters that are passed to the central relational database manager to be used when retrieving. This information to locate the files on specific sectors of the disk is supplied to the System Manager by the Archive Subsystem, and this information is added to the database records keyed to that file control number.

The Archive server also accepts requests for image data over the multibus. The requests contain information including the disk number and sector where the image data are to be found. The subsystem then retrieves the data from the proper disk, either in the jukebox or in one of the stand-alone drives. Image data retrieved from the optical disks can be sent over the multibus to either the Workstation or the Print Subsystems.

Archive Subsystem Configuration

The optical storage system consists of one Sony WDA-3000-10 writable disk autochanger (jukebox) with one internal drive controller, and two internal optical disk drives. This system is daisy chained to an external controller and two drives. The two external drives were used to write the image data onto optical disks and to create backup security disks instead of tying up the jukebox with these activities. Both the drives and controllers are themselves controlled over a small computer systems interface (SCSI). The SCSI bus carries all the information to and from the writable disk controller and jukebox. The SCSI interface includes all the commands necessary for complete control of these devices. The interface between the Sony controller and the two external drives that it controls is a Sony communications bus specially designed for their writable disk systems. Although all the drives in the system can write data onto any disk used in the ODISS system, only the external drives actually do the recording of image data during system operation.

When all the files of a particular block have successfully passed the quality control function, the image data for that block are written to the optical disk. This function interfaces with the System Manager Subsystem to get information concerning where on the disk the Archive Subsystem should start recording a specific file. The System Manager Subsystem keeps the location of a specific file as part of its database management tasks. This relieves the VRTX system from this task. VRTX is the operating system installed on the HK69 where real-time response is of critical importance or where only a few support modules are required. VRTX is an acronym for Versatile Real Time Executive, and is considered "real time" because it has an efficient multitasking preemptive priority scheduling system based on interrupts. It is especially well suited to ODISS because it can indirectly handle many of the monitoring functions of the request network's layer 2. A file can be in any of three locations: on a magnetic disk in the capture storage element; on an optical disk mounted in one of the external drives; or on one of the optical disks within the Sony jukebox. The information

needed to identify the file location, either on magnetic or optical media, is part of the database system. The Archive Subsystem has its own disk mapping procedures.

Requests for digital images originate at the workstations. The requests are sent to the System Manager Subsystem. The file control number of the image data is either received from the workstation which generated the request or is retrieved from the index database. The System Manager routes the request to the Initiation and Monitoring Subsystem. This request is passed to the Archive Subsystem over the multibus interconnect if the image data to be retrieved resides there. The request is now serviced by the Archive Subsystem. The image data requested are read by one of the writable disk drives and sent back to the Heurikon over the SCSI bus. This information is sent to any of the three workstation server elements, or to the printer server elements, as appropriate. The transfer of the image data to these other subsystems occurs over the multibus.

Archive Server

The archive server is the HK68/M10 in the core system. This board is interfaced directly to the writable disk controller mounted within the core hardware enclosure. This interface is implemented off the SCSI port of the Heurikon computer. A communications link is completed with signal cable, and there also exists a direct communication line with the jukebox through a signal cable daisy chain from the controller to the jukebox.

B.8 Staff Retrieval

ODISS includes two workstations for staff retrievals. The terminals were intended primarily for gathering data about the feasibility of having NARA staff perform CMSR searches using ODISS to reply to mail in genealogical inquiries.

B.8.1 Station Workflow

After logging on, staff members see the image viewing area on the left side and menus on the right side of the display screen for completing the CMSR index fields to construct a search (see Figure B-15). The staff member fills as many of the following CMSR search fields as are known from the information in the mail in request: last name, first name, middle name, and then code values for rank in, rank out, regiment, and up to three companies. When information for fields is unknown, the fields are left blank, and the search is made on only those fields that are known.

After the fields are completed, the search is begun by pressing a function key. If a single file matches the search parameters, the file control number, index information for all the index fields, and the number of images in the file is displayed on the screen. When there is no match, the system returns a message indicating nothing was found. If there are several files that match the search parameters, the results are listed on the screen as candidates. The list can be scrolled and individual files highlighted by using the arrow keys; and as the list is scrolled, full index and file information is displayed for each file as it is highlighted (see Figure B-16).

The file images are retrieved for viewing by the use of function keys (see Figure B-17). The PAGE UP and PAGE DOWN keys are used to move forward and backward through a file one image at a time in sequential order. Different function keys activate image rotation, image

CMSR Search Screen

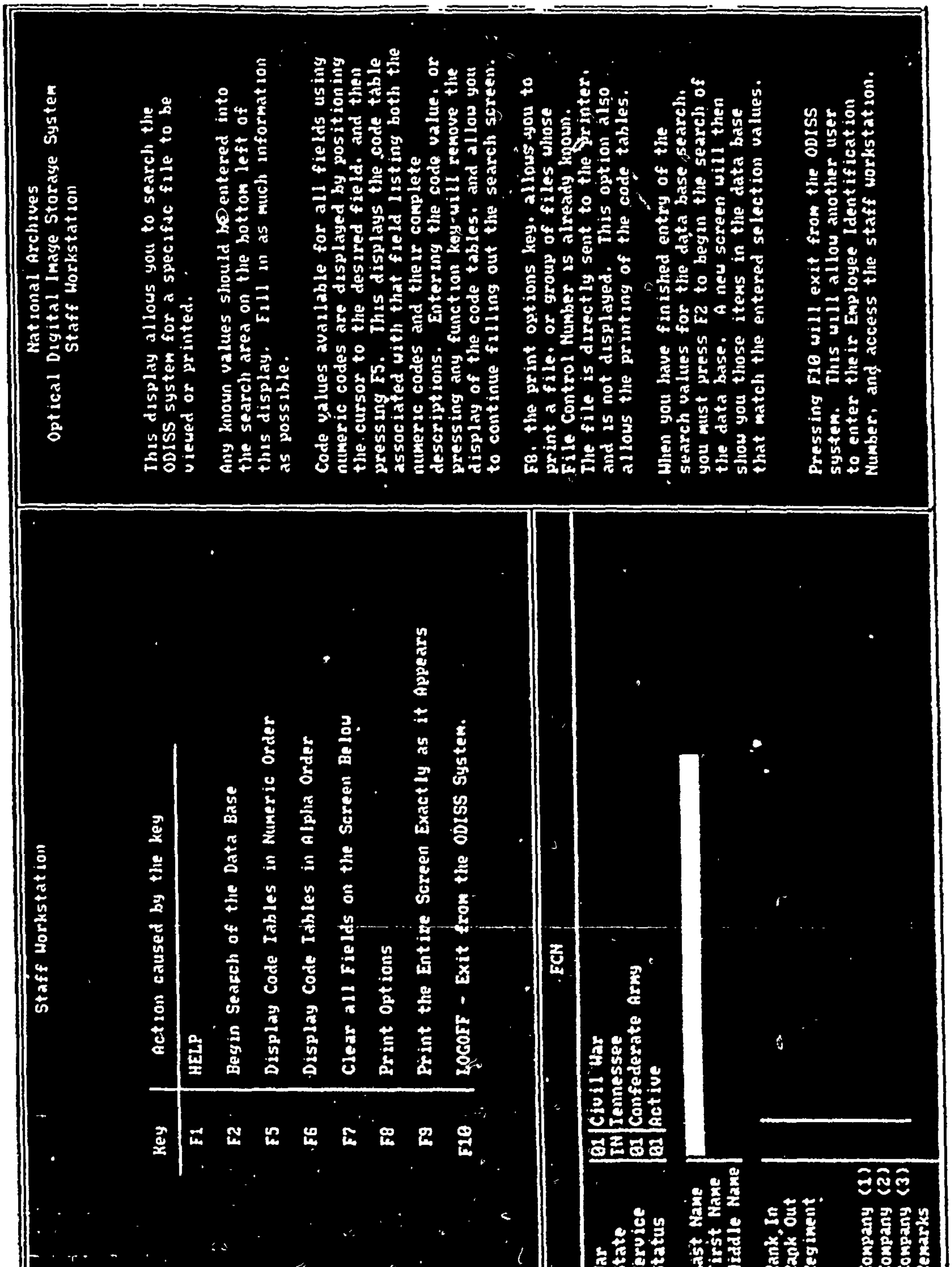


Figure B-15

Highlighted File

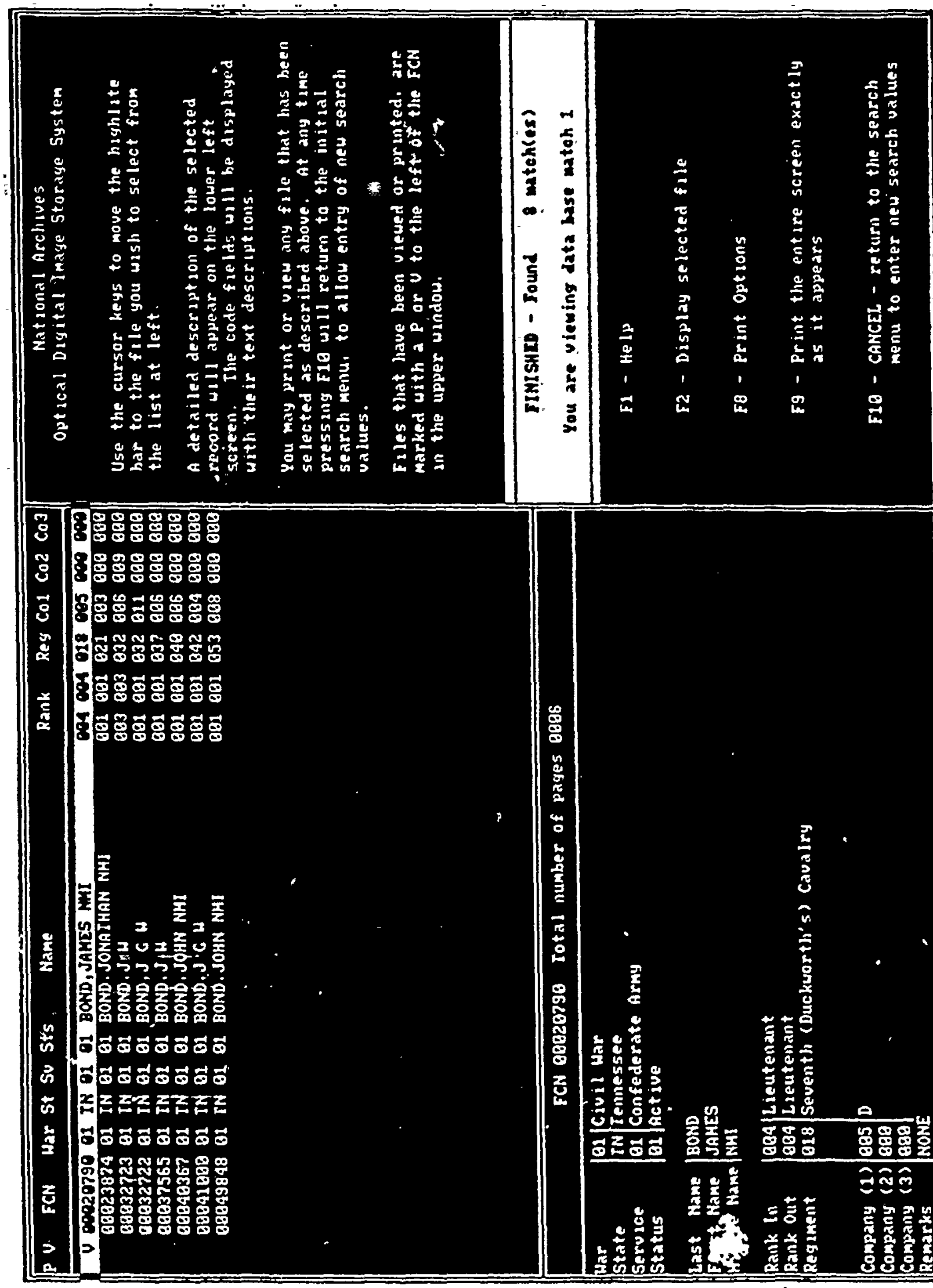


Figure B-16

Displayed Image

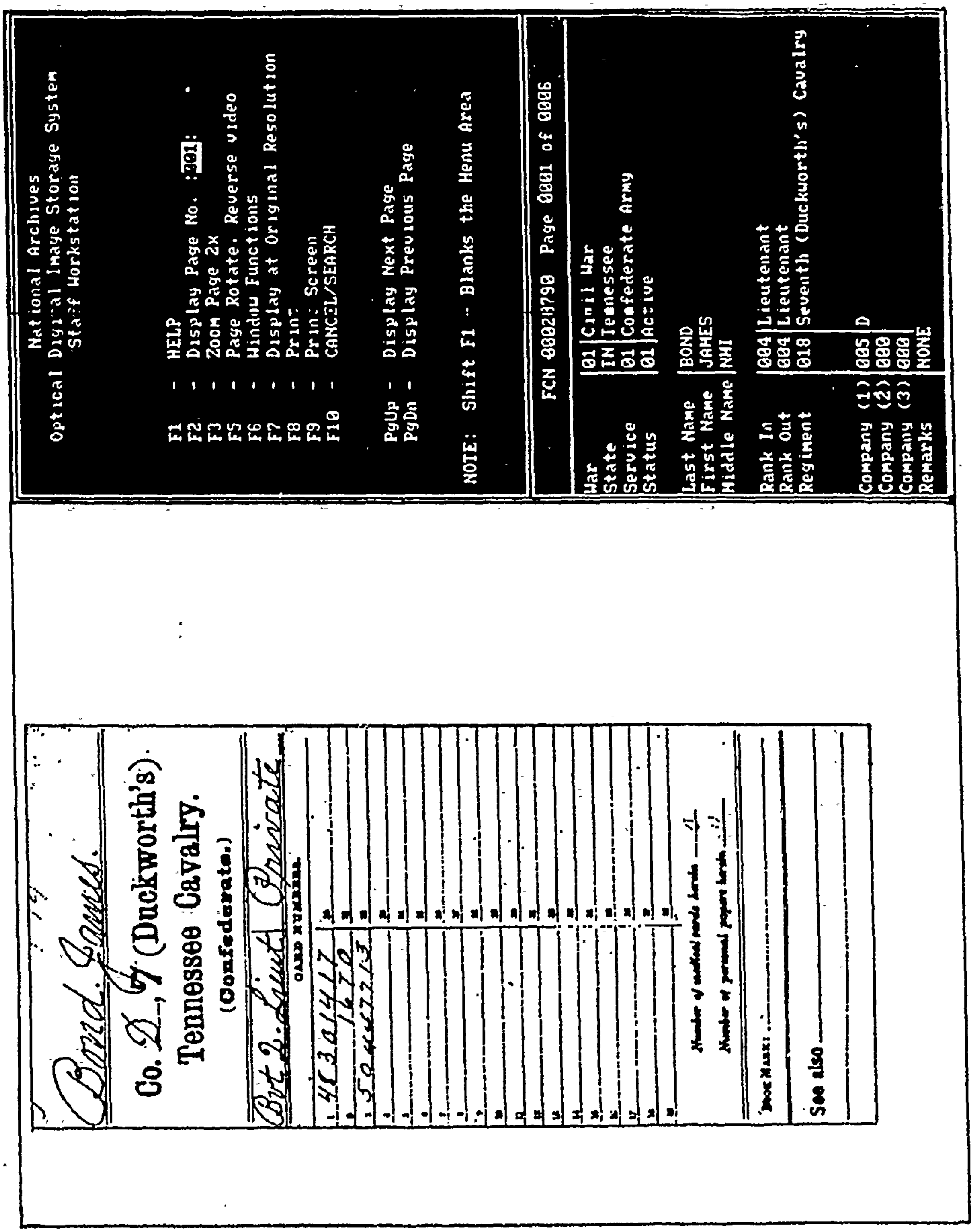


Figure B-17

zoom, movement directly to a numbered image out of sequence, and the various printing options.

The print options include printing the hit list of possible files (see Figure B-18), all the images in a file, or designated images within a file. When prints are made, the system calculates the cost of the copies and produces a cover sheet that lists the file control number, the number of pages printed, and the cost of the copies. The print choices include a batch mode that allows NARA staff to gather into one group a number of paid orders for copies of files that can be printed in a single operation.

B.8.2 Hardware Configuration and Software Capabilities

In both hardware and software, the staff workstations are basically the same as the indexing workstations described earlier. The hardware consists of the same Sperry (Unisys) IT CPU, Discorp image processing board, and video display monitor as well as RS-232 and RS-422 signal cable connections to Heurikon processors in the core cabinet that are the basic elements in all the workstations.

Unisys's software controls the retrieval menus and functionality. The CMSR database that is accessed at retrieval is built under the UNIFY database management program. The software also provides the modularity that permits any number of workstations to be employed for staff retrieval.

B.8.3 Non-CMSR Files

Although non-CMSR files created at the high speed, low speed, or microfilm scanners cannot be accessed at the index or quality control workstations, they can be retrieved and viewed at the retrieval workstation. The non-CMSR side of the retrieval function is accessed during log-on. After entering one's identification number and password, a user is given the choice between selecting CMSR or non-CMSR records.

Once the non-CMSR alternative is chosen and retrieval is selected, the retrieval workflow and capabilities are essentially the same as for CMSR files. Also, like CMSR records, file control numbers are automatically assigned to non-CMSR files by the system during scanning; and, if known, they can be used to retrieve non-CMSR index records and files.

The index records for non-CMSR files are in a database that, like the CMSR database, is established under UNIFY. The major difference is the limited number of index fields available for non-CMSR files. There are only two fields; they are alphanumeric fields for short and long descriptions of the files.

B.9 Public Retrieval

ODISS includes a workstation for the public. The intention was to provide a workstation where the general public could follow screen instructions to teach themselves how to conduct searches for Tennessee CMSR files. The station was placed in the Microfilm Reading Room, which is dedicated to self service use of microfilmed records. It was intended that people using ODISS would conduct research on their own, with a minimum of NARA staff assistance.

Image Search Results

```
Dec 18 1989                    NATIONAL ARCHIVES
                      OPTICAL DIGITAL IMAGE STORAGE SYSTEM
                              INDEX SEARCH RESULTS

    FCN           NAME        RECORD SUMMARY

00020790  BOND, JAMES NMI
          WAR - Civil War  STATE - Tennessee  SERVICE - Confederate Army
          STATUS - Active  RANK IN - Lieutenant  RANK OUT - Lieutenant
          REGIMENT - Seventh (Duckworth's) Cavalry  COMPANY 1 - D
          REMARKS - NONE

00023874  BOND, JONATHAN NMI
          WAR - Civil War  STATE - Tennessee  SERVICE - Confederate Army
          STATUS - Active  RANK IN - Private  RANK OUT - Private
          REGIMENT - Eighth (Smith's) Cavalry  COMPANY 1 - B
          REMARKS - NONE

00032723  BOND, J W
          WAR - Civil War  STATE - Tennessee  SERVICE - Confederate Army
          STATUS - Active  RANK IN - Sergeant  RANK OUT - Sergeant
          REGIMENT - Twelfth (Green's) Cavalry  COMPANY 1 - E
          COMPANY 2 - H  REMARKS - NONE

00032722  BOND, J G W
          WAR - Civil War  STATE - Tennessee  SERVICE - Confederate Army
          STATUS - Active  RANK IN - Private  RANK OUT - Private
          REGIMENT - Twelfth (Green's) Cavalry  COMPANY 1 - K
          REMARKS - NONE

00037565  BOND, J W
          WAR - Civil War  STATE - Tennessee  SERVICE - Confederate Army
          STATUS - Active  RANK IN - Private  RANK OUT - Private
          REGIMENT - Fourteenth (Neely's) Cavalry  COMPANY 1 - E
          REMARKS - NONE

00040367  BOND, JOHN NMI
          WAR - Civil War  STATE - Tennessee  SERVICE - Confederate Army
          STATUS - Active  RANK IN - Private  RANK OUT - Private
          REGIMENT - Fifteenth (Stewart's) Cavalry  COMPANY 1 - E
          REMARKS - NONE

00041000  BOND, J G W
          WAR - Civil War  STATE - Tennessee  SERVICE - Confederate Army
          STATUS - Active  RANK IN - Private  RANK OUT - Private
          REGIMENT - Sixteenth (Logwood's) Cavalry  COMPANY 1 - C
          REMARKS - NONE

00049848  BOND, JOHN NMI
          WAR - Civil War  STATE - Tennessee  SERVICE - Confederate Army
          STATUS - Active  RANK IN - Private  RANK OUT - Private
          REGIMENT - Twenty-first (Wilson's) Cavalry  COMPANY 1 - G
          REMARKS - NONE
```

Figure B-18

B.9. Station Workflow

Workstation display screens are designed to guide the general public in the use of function keys, code tables, etc., to construct searches, to retrieve files, and to print index lists and files' images.

Learning from these on-screen instructions, the public fills in the following search fields: last name, first name, middle name, and then code values for rank in, rank out, regiment, and up to three companies. If the information is unknown, fields can be left blank, and the search can be made on only those fields that are known.

After the fields are completed, the search begins by pressing a function key. If a single file matches the search parameters, the file control number, index information for all the index fields, and the number of images in the file are displayed. When there is no match, the system returns a message indicating nothing was found. If there are several files that match the filled in search fields, the screen displays a list of possible files. The list can be scrolled and individual files highlighted by using the arrow keys, and, as the list is scrolled, full index and file information is displayed for each file as it is highlighted.

The file images are retrieved for viewing by function keys. The PAGE UP and PAGE DOWN keys are used to move forward and backward through a file one image at a time in sequential order. The figures in the previous section on staff retrieval also illustrate the basic screens and steps in public searches since the two processes are similar.

Different function keys activate image rotation, image zoom, movement directly to a numbered image out of sequence, and the various printing options. Image rotation is used to view documents that were scanned sideways or upside down due to their size or other circumstances. Figure B-19 shows the back of a card captured in the high speed scanner's two-sided mode; the back has writing upside down. The image rotation function was used to turn the card right side up, as shown in Figure B-20. In another example, a large requisition document was scanned sideways (Figure B-21), and image rotation was employed for a 90 degree counter clockwise turn to make the requisition easier to read (Figure B-22).

ODISS offers capabilities to increase image size on the display screen. The basic image display mode is 150 dots per inch, while most of the paper records were scanned into the system at 200 dpi and some were scanned at 300 dpi or 400 dpi. By using function key F7, the researcher can display the document at its original scan resolution. Figure B-23 shows a document as it appears on the screen at 150 dpi, and Figure B-24 shows the same document as it appears on the screen at its original resolution of 200 dpi. For even larger display of documents, there is a 2x zoom mode. In the zoom mode, only part of the document can be seen on the screen at one time, and the other portions are viewed by scrolling. Figure B-25 illustrates the screen display of the zoom mode for the same document shown in the previous two figures.

The print options include printing the list of possible files, all the images in a file, or designated images within a file. When the public decides to print, the system notifies them of the copy cost and allows them to choose between stopping or continuing. A laser printer next to the Microfilm Reading Room workstation produces the hard copies.

Upside Down Image

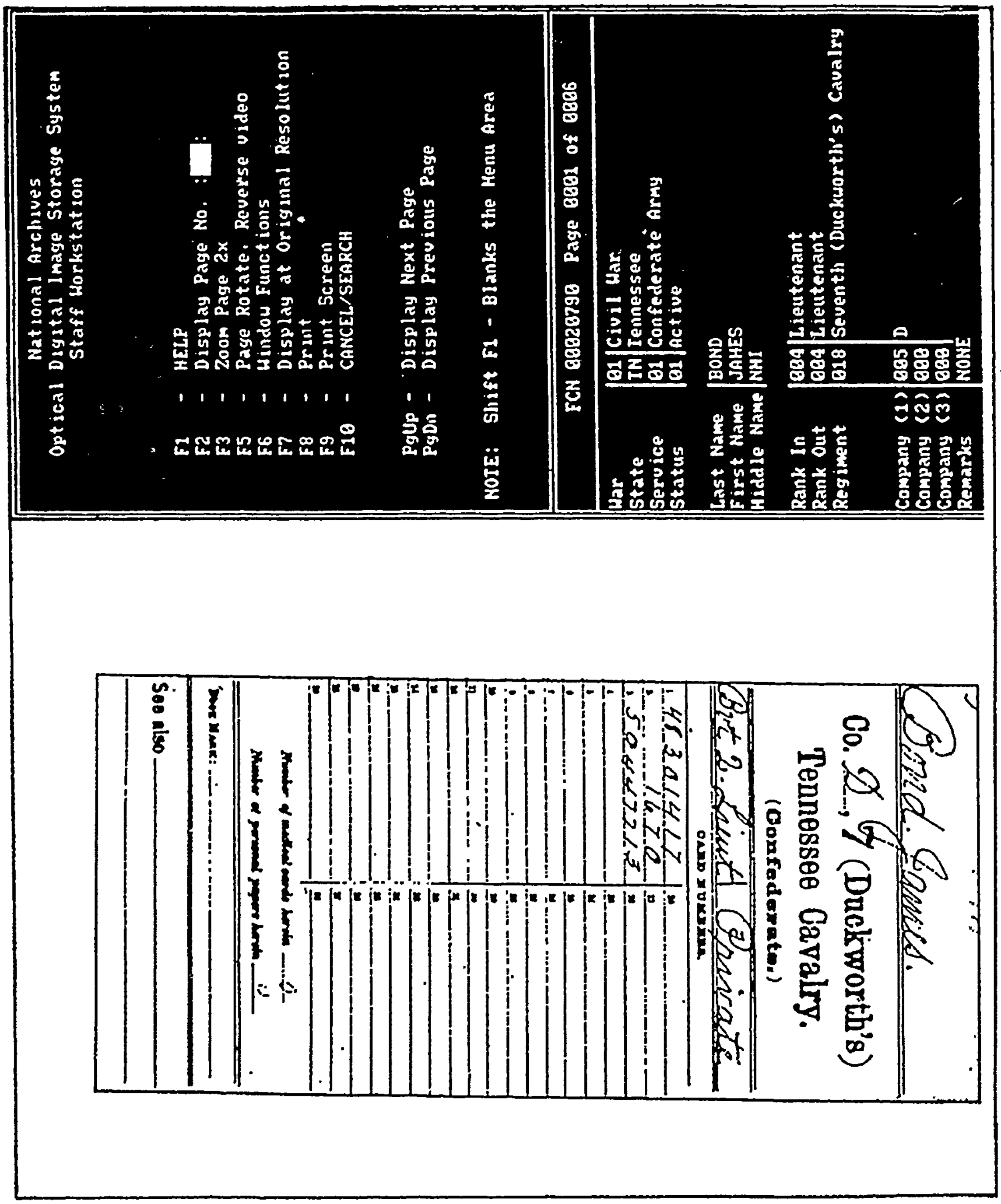


Figure B-19

180 Degree Rotation

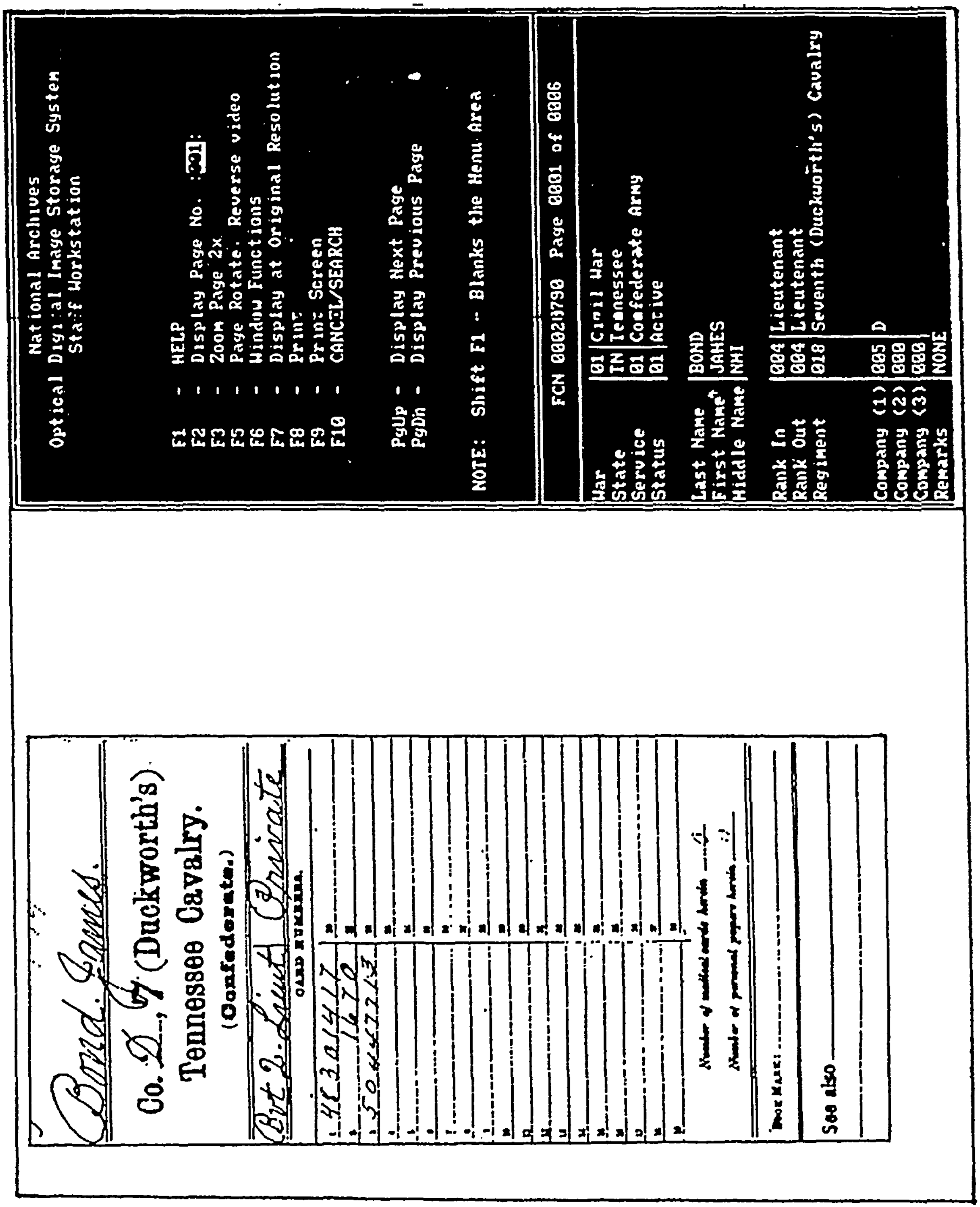


Figure B-20

Sideways Image

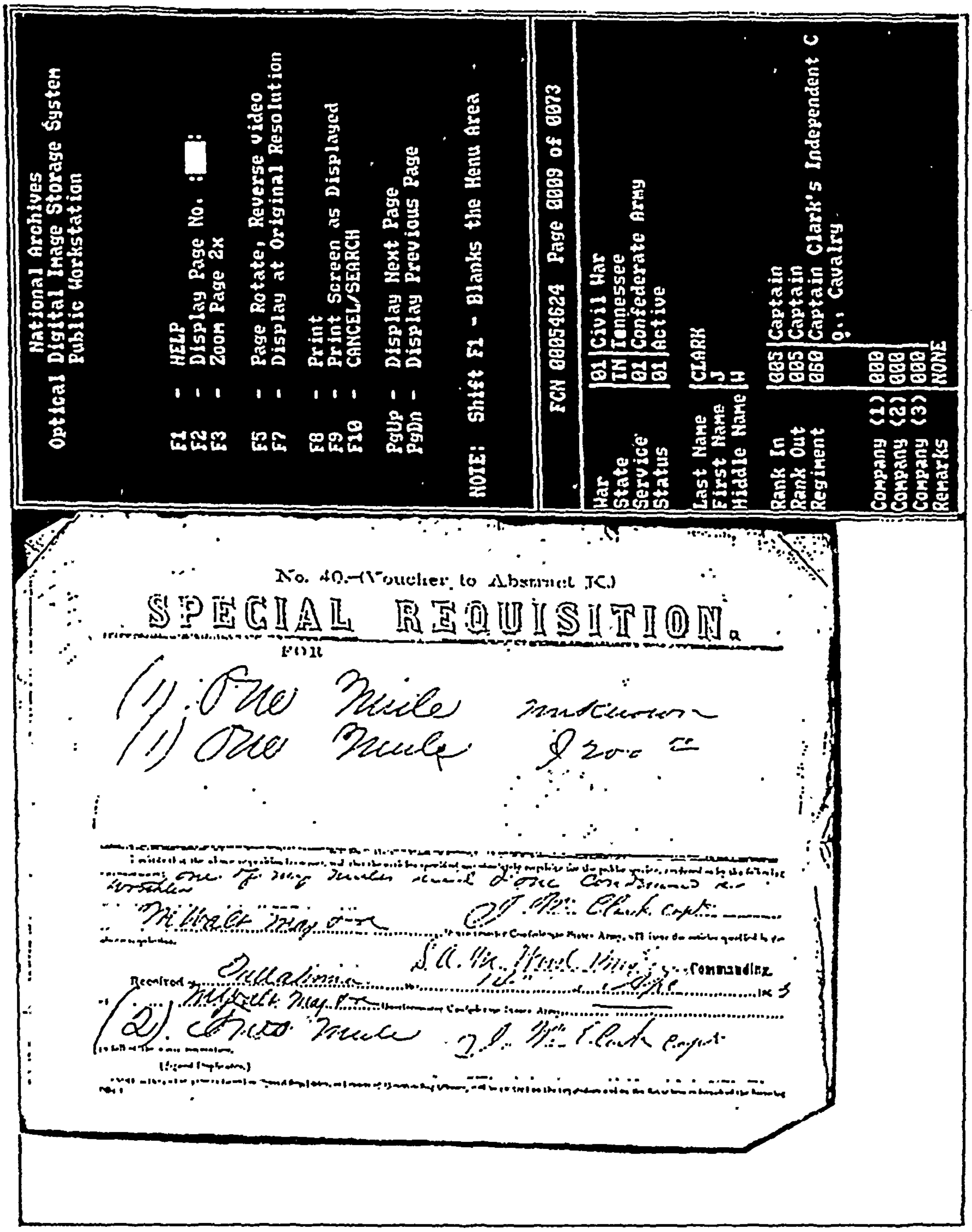


Figure B-21

90 Degree Rotation

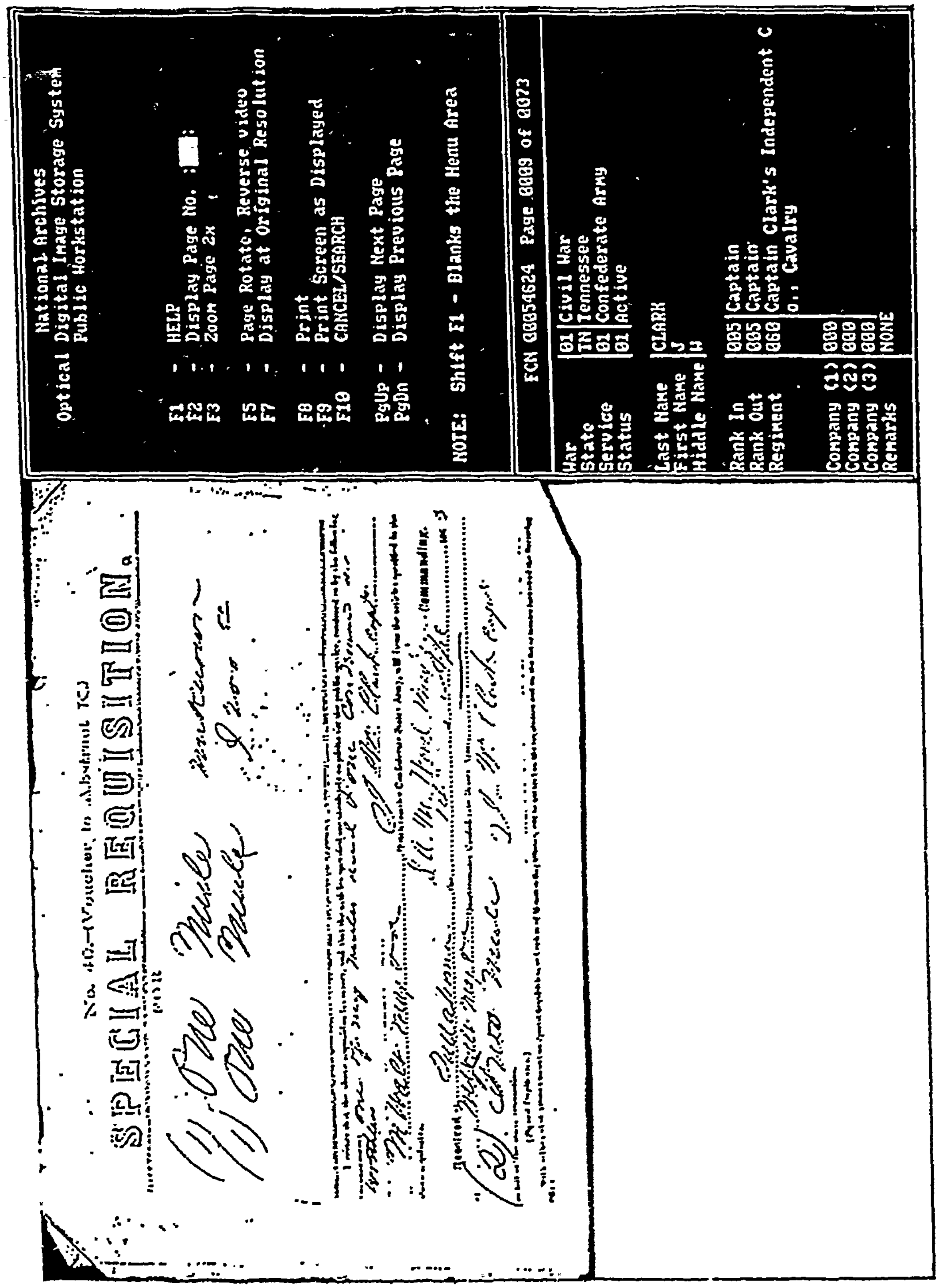


Figure B-22

150 DPI Image

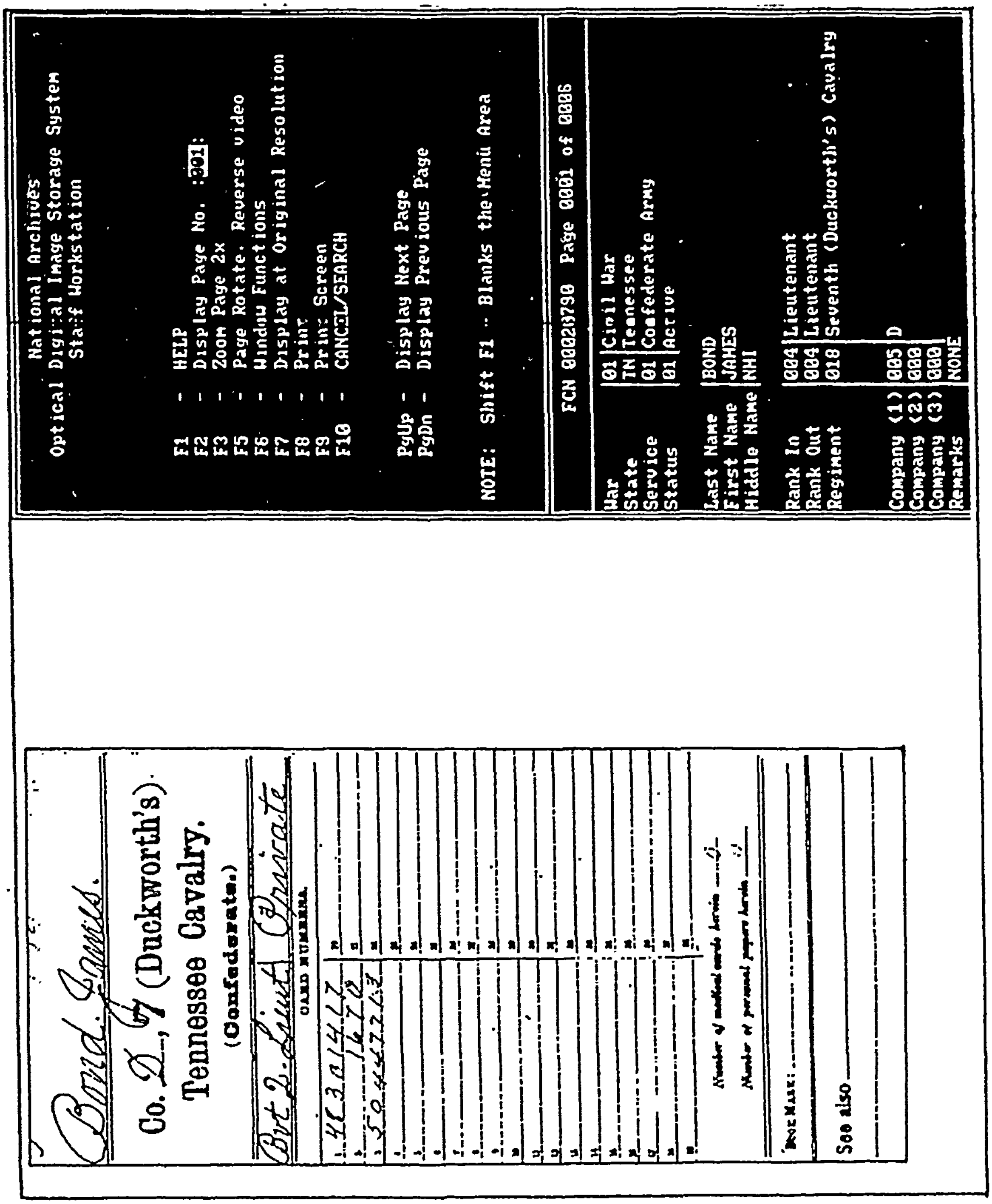


Figure B-23

200 DPI Image

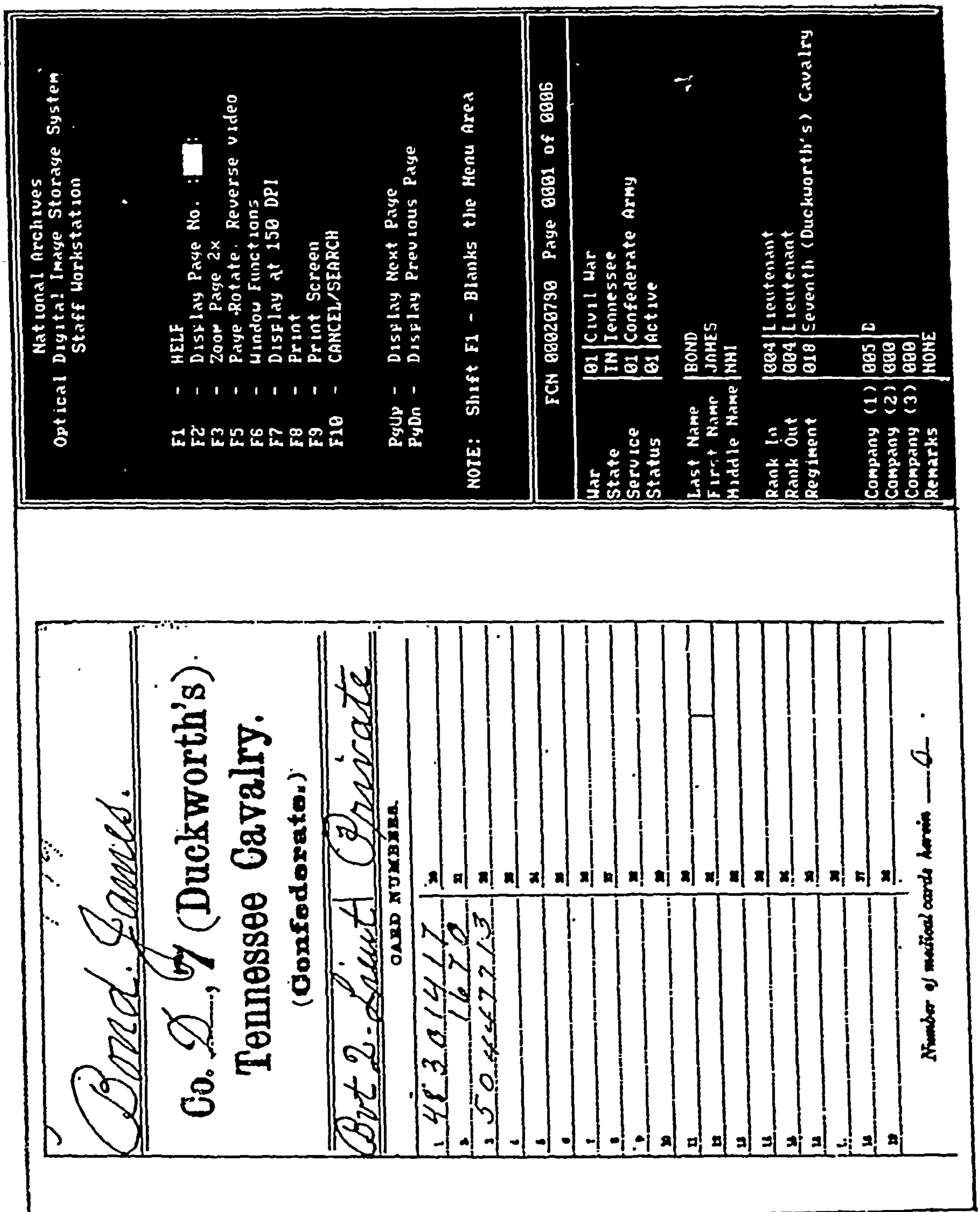


Figure B-24

Zoom Mode

Bond, James.

Co. D, 7 (Duckworth's)

Tennessee Cavalry.

(Confederate.)

Bvt 2 Lieut | Private

CARD NUMBERS.

1	48301417	20	
2	1670	21	
3	50447713	22	
4		23	
5		24	
6		25	
7		26	
8		27	
9		28	
10		29	
11		30	
12		31	

Figure B-25

B.9.2 Hardware Configuration And Software Capabilities

The public workstation includes the same Sperry IT CPU, Discorp image processing board and monitor as well as RS-232 and RS-422 signal cable connections to Heurikon processors in the Core cabinet that are basic workstation elements.

The software also is similar to the other workstations. It includes Unisys written code for the retrieval menus and functionality. The CMSR database that is accessed at retrieval is built under the UNIFY database management program. More detailed hardware and software descriptions are provided in the indexing workstation section.

B.9.3 Adequacy of Screen Instructions

The public workstation in the Microfilm Reading Room was never put into self-service for the general public. This was primarily because the on-screen instructions for the public were determined to be not particularly well-suited for the untrained public. The instructions were first introduced to some NARA staff members who already were computer literate, and these people found the directions to be a bit too complicated for a self-teaching procedure. Consequently, data from a significant cross section of self-taught members of the general public were not obtainable.

B.10 Remote Workstation

An index-only remote site ODISS workstation was installed in the Tennessee State Library and Archives in Nashville, Tennessee. The following sections describe the remote terminal's system design, hardware configuration, and operation.

B.10.1 Configuration

The remote workstation is linked to the ODISS System Manager using a telecommunications modem which transmits and receives data through public phone lines. See Figure B-26 for an illustration of the remote station linkup to the system manager. The PC/IT performs some of the same duties as the SDC-2000 intelligent workstations, namely the communication with the System Manager database information. This data link is the same link provided to the on-site image workstations.

B.10.2 Operation

The Nashville remote workstation first telephones the ODISS facility, where the communications link is established with the "tty2" device driver of the UNIX operating system. Once the link is established and proper entrance codes are transmitted to UNIX, the data communications are brought into a direct link with the SQL interpreter. This link allows the remote workstations to be provided with the same link to the system as for an on-site user. The workstation operates normally, generating searches, and receiving the desired information from the database. The user is then able to use this information to locate the desired files from among the microfilmed copies of the records located at the Tennessee State Archives. The remote workstation also has the ability to generate a print request on a specific file control number received from the System Manager. This request is handled like the requests from the workstations, resulting in the ODISS laser printers producing a hardcopy of the requested images. The prints would then be forwarded to the requestor in Nashville.

Remote Link

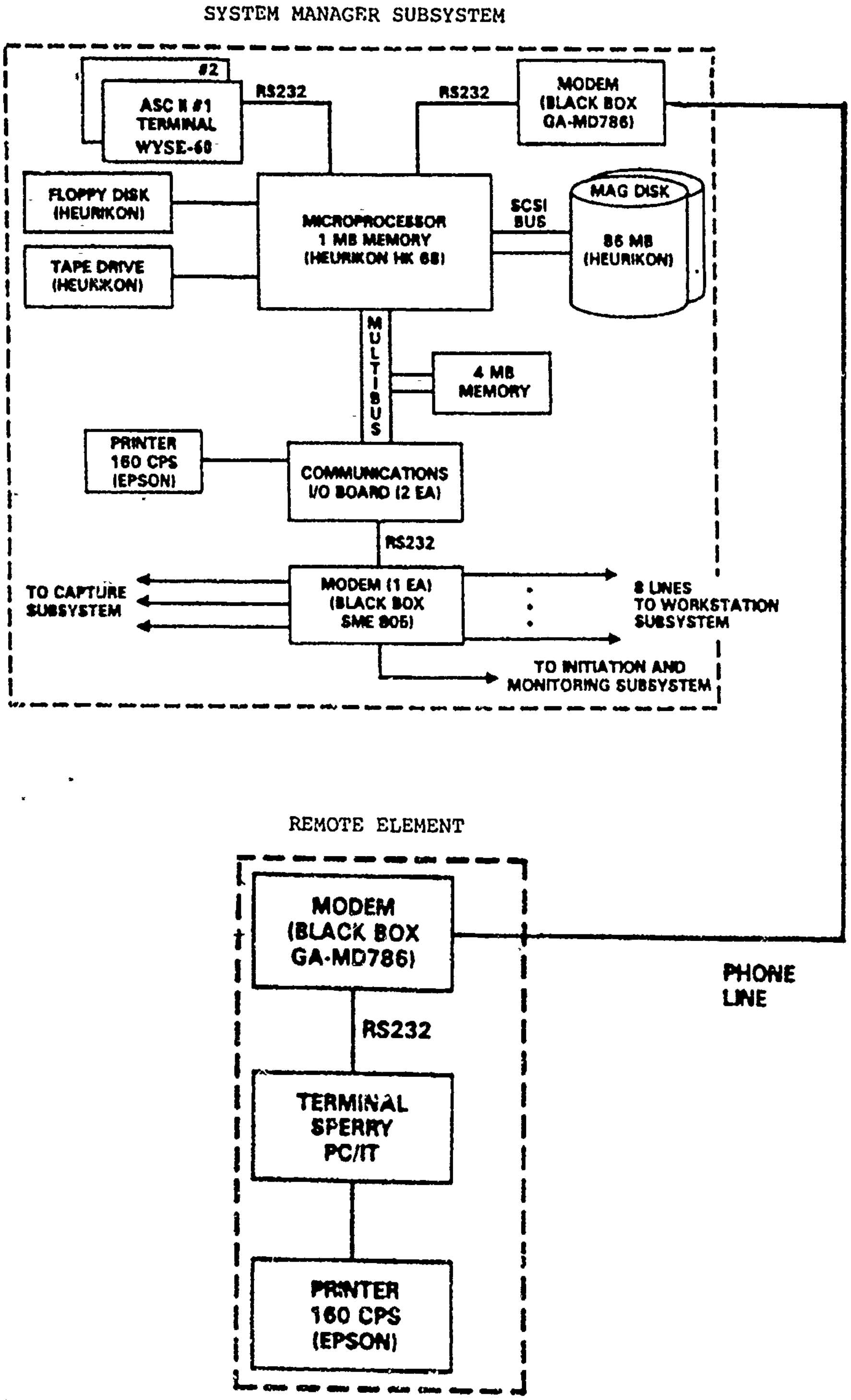


Figure B-26

The remote station is report-oriented in that the Unisys PC performs database inquiries, determines accessibility of records, and gathers information for Tennessee State Archives personnel to use in submitting print requests to NARA.

A remote user would key-enter his known information and request a search of the database. If the search were unsuccessful, the user would then reexamine the parameters entered. The computer is unable to compensate for spelling differences, but the use of a wild card is available. ODISS should be compared to the use of an enhanced library card system, one in which the documents sought by the user are automatically searched for in the entire records storage area. The user must provide enough accurate detail to the system to allow success. See Figure B-27 for a sample remote terminal search screen which shows the results of an index search which produced multiple findings.

B.10.3 Hardware

The remote system consists of a Unisys PC/IT-1 microcomputer, an Epson printer, and a Black Box modem. The PC/IT workstation is based on the 80286 microprocessor. The remote station has a subsystem board, a monochrome monitor, and a 20 MB hard card (disk mounted on a disk controller board). These micros can run at speeds ranging from 6 to 8 Mhz. The remote terminal has an 80287 floating point math processor which performs as a dedicated numerical processor. The system board supports 9 slots conforming to the IBM PC or AT open-bus architecture, and can use hardware designed for this standard. The remote PC/IT also has a keylock power supply override to increase data security. The 200 watt power supply within the PC/IT can operate with 120 or 240 VAC input electrical service.

<u>Telecommunications Modem</u>

The Black Box 2400+ stand-alone telecommunications modem is designed for public telephone data communications. This modem automatically provides protocol necessary for data communication with other modems over phone lines. The modem features auto dialing, asynchronous or synchronous communication, and full duplex operation. The 2400+ modem converts digital bipolar (NRZ coded) signals to frequency modulated signals of the same information content. The modem also supports an RS-232 line on one connector and a four-wire phone tap on another. This modem translates the RS-232 and phone line information back and forth to enable communication in both directions, at either 1200 or 2400 baud rate.

<u>FX286 Dot Matrix Printer</u>

This printer is a dot matrix printer that prints 200 characters per second in draft mode. It also has near-letter quality at slower speeds. The printer is controlled by the RS-232 interface, using 9-wire print head configuration. Near-letter quality is achieved by striking the character twice at slightly different positions. This device also can print different fonts.

B.11 Laser Printer Subsystem

The workstations can produce hardcopy output using ODISS laser printing functions. Image data and index information can be printed using three Ricoh LP5400, 400 DPI laser printers. Two printers are in the main room, the third is located in Room 400.

Remote Search Screen

```
                 Remote Workstation                     National Archives
   FCN          Name             Rnk Rnk Reg Co  Co          ODISS System
00001113 MCCOMBS,ELMO JOHN       001 002 044 001 002
00002054 MCCOMBS,GEORGE B.       001 002 003 004 005  Use the cursor keys to
00002053 MCCOMBS,GILBERT B.      002 003 005 005 006  select a file from the list
00001110 MCCOMBS,MALCOLM STUA    001 003 005 003 000  the left.  Press F6 to
00002094 MCCOMBS,STEPHEN LEE     001 002 003 001 002  PRINT in detail these search
                                                      results.

           FCN 00002094   Total Pages 0016
War            01|Civil War
State          TN|Tennesee                             FINISHED  Found     5 match
Service        01|Confederate Army
Status         01|Active                               Viewing data base match 5
Last Name      MCCOMBS
First Name     STEPHEN                                 CANCEL allows a new search
Middle Name    LEE                                     to begin.
Rank In        001|Private
Rank Out       002|Corporal
Regiment       003|First (McNairy's) Battalion, Cav         F8     Print
Company (1)    001|Field & Staff
Company (2)    002|A                                        F9     Print Screen
Company (3)    003|B
Note-          KILLED IN ACTION                             F10    CANCEL
```

Figure B-27

B.11.1 Hardware Configuration and Software Capabilities

The printers are controlled by SDC-2100 printer controllers. The printer controllers are interfaced to an HK68/M10 single board computer, which handles the print service requests and sends data to appropriate printers. The HK68/10 has a Maxtor XT-3170 magnetic disk drive that provides 170 MB (unformatted) of temporary storage for the printing systems. Figure B-28 is a block diagram showing how the printers are connected to the system.

There are two SDC-2100 print controller configurations, one acting as the controller for the two main printers, and the second controlling the public printer. The Ricoh LP5400 printers are each interfaced to the SDC-2100 printer controller by means of a 36 conductor interface cable. Eight of the signals transmitted on these lines are bipolar, while nine conform to the RS422 differential signal electrical interface. These cables transmit both image data and control information to the LP5400 printers. Status information is passed back to the printer controllers over this same interface. Data are passed serially at a rate of up to 9 Mhz. Specifications on the LP5400 are:

Laser Printer Specifications

Page size:	8.5 X 11 and 8.5 X 14 inches
Print density:	400 DPI
Print rate:	20 pages per minute from buffer
Video I/O:	RS422
Rate:	9 MHz
Channels:	2 1-bit
Byte rate:	2.25 MB per second
Page rate:	72 pages per minute (burst)
Buffers:	Two alternate read and write
Control I/O:	RS422
Rate:	9600 Baud
Mode:	Full Duplex
Protocol:	RS232
Style:	Desk top

Table B-3

B.11.2 Operation

The Ricoh printers can output up to 30 sheets of printed paper per minute, although the image data rates are less due to the K Byte volumes. The print resolution of 400 DPI provides high quality output for standard text or two-tone image material. The LP-5400 can feed paper from any of its three trays. The two smaller trays holding 8.5 X 11 and 8.5 X 14 inch papers are side mounted. The third tray is a mass paper tray located inside the printer's base. The laser within the printer is controlled by the print engine electronics. The print engine interprets the video signal in order to modulate the printer's laser light output. The laser is "on" to represent a dark dot, and "off" to represent a light dot. The laser light falls upon a metal drum where the photoelectric effect causes a static electric charge to be

Printer Subsystem

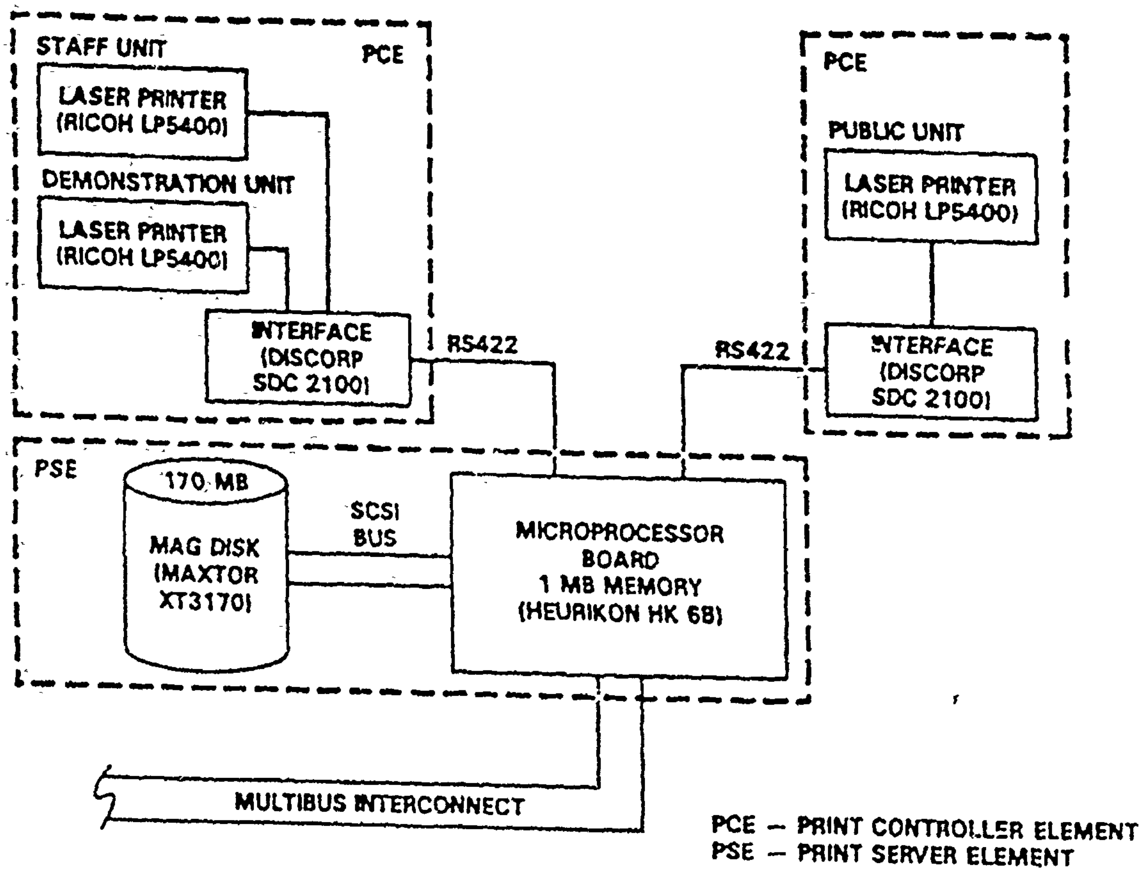


Figure B-28

created on that portion of the drum. Toner dust adheres to these charged sections. Paper pressed against the drum picks up this toner, and heat applied to the paper permanently fuses the toner creating a black dot on the paper.

B.12 System Manager

The responsibilities of the individual performing system manager functions are principally, but not solely, related to the operation of three specialized terminals located in the system manager area. Because the System Manager, CSE/ARS, and Archive Control terminals control day-to-day allocation of material in process, collection of statistical data, and transfer of document images and indexing information from magnetic to optical disk, successful operation of these terminals is vital to the smooth performance of ODISS. In addition to performance of terminal-related tasks, the system manager also oversees certain other aspects of the system.

B.12.1 Hardware Configuration

The system manager station consists of three terminals: the System Manager terminal, the CSE/ARS terminal, and the IMS/Archive Control terminal.

System Manager Terminal

Hardware for the System Manager terminal consists of a Wyse-60 ASCII terminal and a 60-dot-per-second Epson printer, model FX-286. Both elements are connected by RS-232 cables to a single-board Heurikon HK68 microcomputer, with 1 Mb local memory, "running a UNIX operating system and a UNIFY database management package."[92] In addition, the System Manager has been provided with two 86 Mb magnetic disks, connected to the HK68 microcomputer via the Small Computer System Interface (SCSI) bus; a floppy disk drive; a Heurikon magnetic disk streamer; and two communication I/O boards. An additional Wyse-60 ASCII terminal is located in the NSZ offices and "connected to the Heurikon over a short-haul modem link."[93]

CSE/ARS (Capture Storage Element/Archives Subsystem) Terminal

Hardware for the CSE/ARS terminal consists of a Wyse-60 ASCII Terminal connected via an RS-232 cable to a Black Box ABC Switch box. The switch box, which permits users to change from CSE-related ("A") functions to ARS-related ("B") functions, is connected by RS-232 cables to the appropriate microprocessor boards. If the Wyse-60 Terminal is not available, a Qume Terminal, model QVT-D1, usually connected to the film scanner, may be used in its place.

IMS/Archive Control Terminal

The Archive Control Terminal is used for two types of system-related actions. Its principal purpose is to initiate processes on the Initiation and Monitor Subsystem (IMS); thus this

[92] *Unisys Optical Digital Image Storage System (ODISS) Volume XI - ODISS Operations Manual*, page 3-16.

[93] *Unisys Optical Digital Image Storage System (ODISS) Volume IV - Hardware Description*, page 94.

terminal is sometimes referred to as the IMS Terminal. On occasion, the hardware provided for the IMS is used to run other ODISS related programs.

The Hardware for the Initiation and Monitor Subsystem consists of a Heurikon M68010 computer board with 1 Mb local memory, a Heurikon magnetic tape streamer, a Floppy disk drive connected to the Heurikon CPU, 86 MB Magdisk connected to the Heurikon CPU, and One Wyse-60 ASCII terminal connected over an RS-232 communication line.

B.12.2 Operations

Different functions are performed at each of the three terminals.

System Manager Terminal

The System Manager terminal maintains and controls data on employees, workstations, and material on magnetic disk. From the System Manager Main Menu, nine basic database functions, Code Table Maintenance, CMSR/NON-CMSR File Maintenance, User Type Maintenance, Employee Profile Maintenance, Workstation Assignment Maintenance, Main Report Menu, Archive Management, Write Database Backup, and Read Database Backup may be entered. In addition, entries following the "SELECTION:" prompt allow for entrance into other screens and menus, as well as the UNIX shell and SQL capabilities.

Code Table Maintenance

Among the options of the System Manager Main Menu is Code Table Maintenance. Code Table Maintenance provides a seven-selection submenu enabling the user to enter tables relating to the codes for war, state, service, status, rank, regiment, and company. Records in Code Table Maintenance, as in the majority of System Manager Main Menu options, may be examined in one of four modes: add, delete, inquire, or modify. In the case of Code Table Maintenance, these options permit the user to add a new code, to modify or delete an existing code, or to conduct an inquiry concerning one code, a specific range of codes, or all codes under control of the particular table. These four modes operate in the same fashion for records available through other System Manager Main Menu and submenu options.

At present, the war, service, and status code tables, available under Code Table Maintenance, have only one code each; the state table has fifty-three codes (one for each of the fifty states as well as the District of Columbia, Puerto Rico, and the Virgin Islands of the United States); the rank table, fourteen codes; the regiment table, 205 codes, all for Tennessee regiments; and the company table, 1965 codes, all for companies of Tennessee regiments. The Code Table Maintenance option has been used principally for the addition of new companies to the company code table and for the modification of regiment titles under the regiment code table.

CMSR/Non-CMSR File Maintenance

Selection of the CMSR/NON-CMSR File Maintenance option on the System Manager Main Menu permits the user to enter screens for both Compiled Military Service Records and Non-CMSR Maintenance. Screens for each of these record types provide the user with relevant index information. Like the Code Table Maintenance options, both the Compiled Military Service Records and the Non-CMSR Maintenance screens provide addition, deletion, inquiry, and modification capabilities. Both screens provide the full text of the index information (with the exception of the remarks field for the CMSR), as well as the equivalent numeric

code. Because the Compiled Military Service Records screen lacks information on the location of archived files on optical disk, in addition to omitting the contents of the remarks field, the "CMSR5" command (refer to page 271) is generally consulted in its place during day-to-day operation of the system.

User Type Maintenance

User Type Maintenance controls the cost of prints. A price may be set for each page printed from a specific workstation element. As staff and public workstations alone are presently capable of initiating prints, only they and an "unknown" workstation have been provided with this feature. The cost per page presently appears as $.10; but no real price values have been assigned. The standard add, delete, inquire, and modify modes are available on this screen.

Employee Profile Maintenance

Employee Profile Maintenance has a submenu which grants access to two screens, Employee Maintenance and Employee/Workstation Id Maintenance. Both screens permit inquiry, addition, deletion, and modification.

The former screen is simply a catalogue of persons authorized to use the system. This feature lists each employee's identification number (EIN); last, middle, and first name; password; the number of files for which the employee may search on the staff retrieval terminal during one database search; and the number of prints the employee may request on the system at one time. Deletion of this record is not possible without prior deletion of all other records relating to the employee, including management reports containing data on that person. Records for employees, newly authorized to use ODISS, may be created at any time under the addition mode.

The second screen available from the Employee Profile Maintenance submenu, Employee/Workstation Id Maintenance, is an account of the workstations individual employees are authorized to use. Employees are identified by EIN, last and first names; workstations by workstation name and equivalent numeric code. Using the addition and deletion modes respectively, employees may either be granted use of additional workstations or have authorization to use currently assigned workstations removed.

Workstation Assignment Maintenance

Choice of the Workstation Assignment Maintenance option provides the user with a submenu listing three options: Terminal Maintenance, Workstation Maintenance, and Workstation/Terminal Maintenance. Terminal Maintenance lists the UNIX tty number for each terminal. Workstation Maintenance contains a record of the numeric code given to each workstation function, such as index or quality control. Workstation/Terminal Maintenance combines the information from the two previous submenus by listing the workstation functions which a specific terminal can perform. Through use of the basic addition, deletion, and modification modes, terminals or workstation functions may be added or deleted from the system, while the workstation functions assigned to specific terminals may be added, deleted, or changed. The inquiry mode is also available under all three submenus.

Main Report Menu

The Main Report Menu, accessible from the System Manager Main Menu, displays three types of reports: Management Accounting Reports, Display Rollup Log, and Employee/Workstation Permissions.

The Display Rollup Log screen displays a listing of the number of additions, counted by EIN log-ons to each workstation element, made to the reports each week.

Employee/Workstation Permissions permits a search of a range of EINs. Such a search yields a catalogue of all individuals authorized to use ODISS, listed by last name, first name, and EIN; and the specific workstations to which each has been granted access.

The Management Accounting Reports option produces reports providing data on the amount of material processed or retrieved by each ODISS element. These reports are available for daily, weekly, quarterly, and yearly time periods. Data in all reports is grouped by EIN with a total for the entire date period included at the end of the report. Prior to producing the report the management report program requests beginning and ending dates for the time period of interest as well as the desired Employee Identification Number (EIN) or range of Employee Identification Numbers. No report can be produced unless the time period which the request covers has been completed. Management accounting reports have the option of either being displayed on the terminal or printed.

All management accounting reports include a statement of the time-period and EINs for which a search was conducted, and the date and time when the report was run. All reports are also broken down by EIN. For weekly, quarterly, and yearly reports data within the EIN section is listed by ending date of the appropriate time-period; for daily reports by each log-on. For all workstation elements, except quality control, all data within each date is further broken down by Unix tty Number. In each report, totals in all categories are given for every EIN; grand totals for all categories appear at the end of the report.

The *High Speed Scanner Management Report* provides the number of images scanned and files processed at the High Speed Scanner.

The *Low Speed Scanner Management Report* yields data on files entered or rescanned at the Low Speed Scanner. For material first entered at this element, the number of images scanned and files processed is provided, for records rescanned, the number of images rescanned and files redone.

The *Index Workstation Management Report* lists number of files indexed.

The *Quality Control Workstation Management Report* shows the number of files reviewed, images reviewed, files approved, files rejected, images rejected, images not scanned (i.e., blank pages inserted), and images designated "best image." A file is counted as "reviewed" each time it is retrieved by a quality control workstation, even if no final disposition is made of the file. Because a file may be reviewed more than once, the total number of files reviewed is, in many instances, greater than the combined totals of the number of files approved and the number of files rejected. This factor should be taken into account when evaluating Quality Control Workstation data.

The *Staff Workstation Management Report* provides the number of searches conducted, searches printed, and images printed.

The *Public Workstation Management Report* also enumerates the number of searches conducted, searches printed, and images printed.

The *Remote Workstation Management Report* registers the number of searches conducted and copies printed. The User Type (i.e., the element authorized to print copies) is also indicated.

Anyone wishing to interpret the management accounting reports must be aware of the specific computation measures used for each report. The discrepancy in the Quality Control Workstation Report between the total number of files reviewed and the combined total of the numbers of files approved and rejected is the most notable example of the need for this requirement.

Archive Management

Archive Management is both the most frequently utilized and the most important selection offered by the System Manager Main Menu. The Archive Management submenu accessed from the System Manager Main Menu offers four screens: Display Status of Previous Archive, Blocks Ready to Archive, Optical Disk Free Space, and Initiate Archive Process.

The Display Status of Previous Archive screen records the result of the last archive initiated. This record indicates date and time of initiation, number of block being archived, number of files in that block, and deletion of block record from the system manager database. For individual files within the block, the display records writing of request to archive, reading of Archive Response Message, update of the CMSR and ODISK files, deletion of the FCNBLOCK record, and number of the individual file out of the total number of files in the block. In addition, Display Status of Previous Archive lists any errors resulting in the failure of an archive to initiate or to finish successfully. Both a brief description of the cause of the archive failure and the numeric code of the error are given in such instances.

The Blocks Ready to Archive screen shows the numbers, in ascending order, of all blocks ready for archive. Because it does not require unnecessary entry of the Initiate Archive Process screen or laborious search for the status of individual blocks, this option is useful for quick and convenient reference.

The Optical Disk Free Space screen indicates the amount of unused space available on each side of an optical disk. All initialized optical disks are listed on this screen by node, volume number, and side.

Write Database Backup

The Write Database Backup option allows the user "to back up the magnetic disk database to floppy disk or cartridge tape."[94] Because any interference will cause the database Backup to abort before completion, it is essential that no other users attempt to accession the

[94] *Unisys Optical Digital Image Storage System (ODISS) Volume XI - ODISS Operations Manual*, page 4-8.

database while this program is running. During normal production the Write database Backup is done daily.

Read Database Backup

The final selection offered by the System Manager Main Menu is Read Database Backup. This program "allows the user to restore the magnetic database from the floppy disk or cartridge tape."[95] As with the Write Database Backup program, no one else may use the database while the backup is being read.

System Manager "SELECTION:" Options

In addition to the nine functions available from the System Manager Main Menu, access to several other key features of the System Manager may be gained on any screen displaying the "SELECTION:" prompt, when appropriate commands are entered following that prompt. Several of the more commonly used commands are worthy of further discussion.

The "BLOCK" command is used to determine the status of a block of material between its creation and its final transfer to optical disk. The block maintenance screen lists block number, block status (open or closed on the system manager), block stage (entry, index, quality control, pre-scan, rescan, pre-archive, or archive), number of files scanned, number of files indexed, database type (CMSR or Non-CMSR), media source (paper or film), scan source (film, high speed, or low speed scanner), block filler, block date (date of block's creation), and block time (time of block's creation). The number of files indexed does not appear for Non-CMSR blocks, and appears for CMSR blocks only after completion of indexing for all files within the block. Once again, the user has the inquire, add, modify, and delete options.

The "FCNBLOCK" command allows the user to determine the status of an individual file in the system. The FCN Block Maintenance screen provides the file control number, block number, sequence number (number of the file within the block), assigned number (number assigned to the file within the block), function block status (open or closed), function block stage (entry, index, quality control, rescan, or archive), and number of images. The sequence and assigned numbers are usually identical. Again the user has the ability to inquire about, add, delete, or modify a record.

The "FCNPAGE" command provides useful data relating to pages which have been electronically marked for rescan. Individual pages are identified by file control number, page number, block number, and file sequence number. Inquiries for records concerning rescan pages may be conducted from the system manager, where addition, modification, or deletion of such records is also possible.

The "CMSR5" command serves a function similar to that provided by the Compiled Military Service Records selection on the CMSR/Non-CMSR File Maintenance Submenu. However, the CMSR5 screen does not display the written equivalents for numeric indexing codes. The CMSR5 screen does display the full text of the remarks field. More importantly, the CMSR5 screen displays "OPTICAL DISK INFORMATION." For each File Control Number, this

[95] *ibid*, 4-41.

section provides cluster number, node ID, side, volume number, sector number, and number of images. The number of images in the file appears on the database as soon as indexing is completed, and is updated as pages are added or deleted. The side, volume number, and sector number yield crucial information needed for recovery from an abnormal archive termination. CMSR5 is the only feature of the System Manager terminal which provides this data. The user has access to the inquire, add, modify, and delete modes in CMSR5.

The "ODISK" command brings up the Odisk Maintenance screen, which shows Odisk Cluster Number, Odisk Node ID, Odisk Side, Odisk Volume Number, and Odisk Next Sector. Odisk Next Sector indicates the next clear sector on an optical disk, an important factor in recovery from an archive failure. The Odisk Maintenance screen is equipped with the ability to inquire, add, modify, and delete.

The "MAINMENU" command accesses the entire UNIFY menu. Care must be used while in the UNIFY menu subsystem so as not to damage the database.[96]

Of eight Mainmenu options only Data Base Design Utilities is used with any frequency. This function has its own submenu with seven possible selections, of which only Add, Drop B-Tree Indexes has been utilized by NARA personnel. Add, Drop B-Tree Indexes allows the user to add, drop, or rebuild the B-Tree (indexing patterns) for either CMSR or FCNBLOCK records, both of which have three indices.

Shell Commands

The sh (shell) command grants the user entrance to the UNIX shell screen. "Sh is a command programming language that executes commands read from a terminal or a file."[97] In ODISS, shell is generally used to run certain special programs, usually related to optical archives or daily database backups. Four of these programs - "NAMELIST," "NUMCHECK," "FCN2LIST," and "FCNLIST" - are used with sufficient frequency that they merit mention.

The "NAMELIST" program retrieves and prints out all file control numbers within a given range, together with the last, first, and middle names from the relevant index files. Only files which have been indexed will appear on this list. "NAMELIST" may be run at any time; however, it is generally used prior to an archive of a CMSR block, in order to verify that all files have been scanned.

"NUMCHECK" is a program which prints out a list of all file control numbers in which discrepancies between the number of images in the "FCNBLOCK" record and the number of images in the "CMSR5" records exist. All Non-CMSR and unindexed CMSR records should appear on this printout. Any indexed CMSR files in this enumeration should be investigated further; if necessary, appropriate modifications should be made to the "FCNBLOCK" or "CMSR5" records on the System Manager. The "NUMCHECK" program should be run occasionally when blocks containing unindexed or problem files from a previous "NUMCHECK" printout become available for archive. Since the commencement of on-site testing, "NUMCHECK" has largely been replaced by the "PRESCAN" program. The

[96] *ibid*, page 4-48.

[97] *Uniplus+ System V User's Manual*, Section 1, 1. Commands P-Z, SH (1), page 1.

"PRESCAN" program, which is run at regular workstations, includes among its functions, comparison of the "FCNBLOCK" and "CMSR5" records on the System Manager with relevant records on the Capture Storage element.

The "FCN2LIST" program provides a printout of all file control numbers within a block, listing assigned and sequence numbers, file stage, and number of images for each file control number, as well as total assigned numbers and actual number of files in the block. This program may be run at any time between the creation of the block and the deletion of the block record from the system manager. Usually, "FCN2LIST" is run immediately prior to an archive, so that a record of a block's status at that time may be retained.

"FCNLIST" prints out a record of all non-archived file control numbers. "FCNLIST" provides the same file information as "FCN2LIST," but for all blocks appearing on the System Manager database. On an "FCNLIST" print-out file controls numbers are listed in ascending numerical order. File control numbers are not specifically divided by block. Nor are the assigned and actual number of files listed after the final file of a block. In fact, no record of the assigned number of files within a block is provided by "FCNLIST." In addition to file data, "FCNLIST" includes a section of block-specific information. This portion of the report shows block number, block stage, number of files scanned, number of files indexed, media type (paper or film), scan source (high speed, low speed or film scanner), and date of block creation.

The Shell capability is also used occasionally for text and visual editing. The Text Editor (ED) permits the user to create and edit text, as well as to print the text to the screen and to store it in a file. The Visual Text Editor (VI) is a full-screen editor while ED is a single-line editor. These functions were not extensively utilized, although they were employed during the first months of operation to correct certain indexing errors.

Structured Query Language

Structured Query Language, commonly known as SQL, is another important function performed at the System Manager terminal. Although used only occasionally, SQL merits mention because of the essential service it provides with its ability to conduct searches of the database for information not readily obtainable through use of the normal menu options. Such searches can yield useful statistical information, such as the number of files and images on an optical disk. SQL commands may be structured so as to provide more detailed, more specific, or more easily usable data. SQL is accessible from both the "SELECTION:" prompt on menu screens and the Shell screen.

CSE/ARS (Capture Storage Element/Archives Storage) Terminal

The CSE/ARS terminal is used for two distinct, essential functions of ODISS. Under CSE, several functions providing information about data stored on magnetic disk are offered. Under ARS, information pertinent to data already transferred to optical disk is made available.

Capture Storage Element (CSE)

The Help command (HELP) under the terminal's CSE function displays the commands used to accomplish the eighteen different options (including HELP) provided by that function.

Although used only occasionally, the Close Transaction command (CT) is nonetheless highly important. CT permits the user to retrieve a file which has been left open on Capture Storage by closing the appropriate transaction. The relevant CPU (i.e., workstation element) ID is entered at the prompt. A message is then returned indicating whether or not the transaction has been closed. The CPU ID may be obtained from the System Activity (SA) function of the CSE terminal.

The List Directory command (DIR) grants the user the option of viewing directories for three drives - 0, 1, and 2 - which together contain data on all file control numbers stored on magnetic disk. Each file control number is divided by three, the remainder - 0, 1, or 2 - then determines in which drive the file control number is stored. List Directory displays file control number, sector, and length data for both the image file (.DAT) and the Page Descriptor Table (.PDT). .DAT and .PDT data for up to twenty-three files are displayed on each screen. Each directory screen provides the user with the option of either bringing up the next page of the directory or terminating the display. Files which have been archived or deleted are replaced by the words "Deleted Entry."

Display Error Log (DL) provides a log of all errors which have occurred on the system since the error log was last purged.

Dump Sector (DS) "allows any single sector on any single disk to be displayed ... in both hexadecimal and ASCII format."[98] Prior to returning this information to the screen, the CSE must receive replies to prompts requesting the drive and sector numbers. Up to twenty-three lines of data are displayed on each Dump Sector screen.

The Initialize Volume command (IV) is invoked periodically to clear off the magnetic disk. Initialization is performed on one drive at a time. Once initialization is performed all data previously on the initialized drive is permanently lost. Users should be careful to enter the IV command only when they have a drive or drives which they desire to initialize. As there is no simple means of escape from this function once invoked, the user may return to the main CSE directory only by initializing a drive. If no drive is designated, drive 0 will automatically be initialized. If the user does not desire to initialize all drives, special care should be taken that the correct drive number is entered prior to return of the initialization request.

The List Directory command (LS) provides the same information as the List Directory (DIR) command but for only a single file control number. LS also shows the drive on which the file data is located. Data for a single file number may, thus, be obtained, quickly and conveniently, without the necessity of paging through a full directory. In the event that data on a specific file is not available, the message, "DAT File Not Found." and/or "PDT File Not Found." will appear on the terminal screen.

The Page Descriptor Table List (PDT) yields data on each page within a specific file. After prompts requesting the drive and sector numbers are answered, the CSE terminal displays the number of pages, the next PDT sector, and "PDT Page Entry Data" for each page. "PDT Page Entry Data" includes data on page number, cluster length, relative sector, page attribute, data block count, and kilobyte block count. The Page Descriptor Table is often

---

[98] "CSE Diagnostic Terminal," 2.6 in memorandum dated October 27, 1988 to Frank Miller from Ryan Stoutenborough.

consulted following abnormal archive terminations. The drive and sector numbers needed to call up the PDT file for a file control number may be obtained through the DIR or LS commands.

Remove File (RM) removes a file from magnetic disk. After the Remove File command is entered at the CSE terminal the operator is asked to provide the file control number and the drive number. Once this information is entered deletion of the appropriate file is initiated. A message indicating successful or unsuccessful deletion of the .DAT and .PDT files is returned upon completion of the process.

The System Activity command (SA) lists all files presently open on the system. The files are identified by CPU (i.e., workstation element) ID, File Control Number, Status, TIME (VRTX tick Count), and "Action Message." "The ACTION MESSAGE either represents the message 'in progress' or, if there is none, the action associated with the first operation on this FCN."[99] Indications of Bus, LCE, and FCE are also given, even when no file is open. The System Activity command is used only rarely, generally in conjunction with the Close Transaction command.

Volume Status (VS) is one of the commands most frequently invoked at the CSE terminal. Volume Status records are maintained for each of the three drives. Upon the answer of a prompt requesting the drive number, a display is returned to the screen indicating what percentage of the drive is full. In addition, Volume Status provides data on Available Clusters, Used Clusters, Contiguous Clusters, Bad Clusters, Disk Volume Type, Sectors/Clusters, Number of Fat Sectors, Fat Entry Size, Directory Size, and Directory Entry Size.

The Zero Boot Block (ZT) is used occasionally to erase the error log. This procedure is generally performed just prior to a system reset or reboot.

Several other features of the CSE terminal are available, but have not been utilized by NARA personnel. These features are Display Counts (DC), Dismount Disk (DIS), Display Waiting Messages (DW), Mount Disk (MD), Monitor On/Off Toggle (MT), and Sync Disk (SD).

Archive Storage (ARS)

When switched to ARS functions, this terminal provides six options relating to optically stored material.

After the completion of archiving on each side of an optical disk, a directory is created on that side. This directory may be viewed by entering the Directory List command, DIR. The Directory List display has twenty-three lines of data, recorded in two columns. Thus, up to forty-six entries are available on each screen. Each entry lists a file control number, its sector location on the disk, and its length. In the case of files which have subsequently been deleted from the disk, "Deleted" appears in place of the file control number. Directories do not exist for material archived to optical disk sides which have not been filled.

The Dump Sectors command (DS) has a function identical to its equivalent on CSE. Entry of the appropriate disk, side, and sector yields an account of the actual data on optical disk.

[99] *ibid*, 2.15.

If a file has been properly archived the file control number together with the veteran's name appears on the right side of the screen. On the left side of the screen the actual numeric codes for the pointer to the next disk sector as well as for the data in the file appears in hex form. Each Dump Sector screen displays up to twenty-three lines of data.

Dump Sectors displays are used, almost invariably, in the process of recovering from an abnormal archive termination. File data is generally examined to determine that the file has been properly archived and that file pointers eventually point to a blank sector correlating with the next blank sector indicated by the "ODISK" record on the System Manager terminal. In the event the next blank sector on the disk is not the sector indicated by "ODISK," the "ODISK" record is corrected to include the proper disk sector. As the pointer data in each dump is given in hex form, a calculator capable of converting hex numbers to their decimal equivalents is necessary for successful examination of optical disk sectors.

The help (HELP) command provides a directory of the six ARS terminal commands, including HELP.

The other three commands available at the ARS Terminal are: Return to HBug via Illegal Address Trap (QUIT), Record Diagnostic Sectors (RDS), and Record Volume Label (RVL).

IMS/Archive Control Terminal

The Archive Control Terminal is used for two types of system-related actions. Its principal purpose is to initiate processes on the Initiation and Monitor Subsystem (IMS); thus this terminal is sometimes referred to as the IMS Terminal. On occasion, the hardware provided for the IMS is used to run other ODISS related programs.

When set up under IMS, the Archive Control terminal features a main Archive Control screen with a common message area comprising the bottom fifth of the screen. IMS is capable of performing eight basic tasks, invoked by entering a two-letter abbreviation after the "OPERATION:" prompt on the main Archive Control screen. Each IMS selection has its own screen. Change of screen does not affect the common message area.

The Create Directory (CD) option creates a directory of file control numbers for each side of an optical disk. This directory provides the sector location for each file control number; it is the same directory which may be viewed with List Directory (DIR) command under the ARS function of the CSE/ARS terminal. The Create Directory screen requests the Volume ID and Side ID for which the directory is to be created. A message indicating the successful completion or unsuccessful termination of this program will appear in the common message area of the terminal. A directory should be created only after a disk side has been completed.

Copy Volume (CV) permits the user to create a duplicate copy of an optical disk. The Copy Volume screen requests Source Volume ID, Source Side ID, Destination Volume ID, and Destination Side ID. Only one side of a disk may be duplicated at a time. If copies of more than one side of a disk are desired, a separate request must be entered for each copy desired, after the completion of the previous request. Notification that the volume copy has been completed will appear in the common message area.

Abnormal termination of a copy before completion requires that the copy be redone from the beginning on a new disk. There is no way for the copy to continue from where it left off on a partially completed disk copy. Due to this factor a volume copy should be initiated only

when no one else is using the system. Leaving the copy running overnight proved to be the most convenient course of action.

The Delete File (DF) program allows the user to delete a previously archived file from optical disk. The Delete Archived File screen requests the Volume and Side IDs, and the File Control Number. The File Control Number must contain at least four digits in order for the program to initiate properly. If the File Control Number designated for deletion contains fewer than four digits, zeros should be added before the first digit. A status message will appear in the common message area upon successful or unsuccessful completion of this program. Because deleted files are permanently lost, users should take care that the proper File Control Number is entered prior to returning the deletion request.

The Dismount Volume (DV) command dismounts an optical disk from the jukebox or one of the stand-alone drives. The Dismount Volume screen asks only for Volume ID. Success or failure of the dismount will be indicated in the common message area. The dismount procedure is one of the most commonly performed functions of the Archive Control Terminal.

The Index Retrieval (IR) option "retrieves the index information stored within each archival file and returns it for storage to an IMS disk file."[100] The Index Retrieval screen requests the user to provide both Volume and Side IDs. The common message area will display the ultimate status of this operation. Index Retrieval is not commonly used.

The Initialize Volume (IV) procedure records the volume label and predefines "bit patterns in the diagnostic tracks."[101] An uninitialized optical disk is placed in one of the stand alone drives, with side A up. The unit (i.e., stand-alone drive) ID is entered on the Initialize Volume screen. A message stating "The assigned Id for the new volume is (Volume Number)" will appear on the screen when the request is entered. The status of the initialization procedure will be displayed in the common message area. This procedure should then be repeated for side B.

Media Diagnosis (MD) "reads the diagnostic tracks on the specified volume/side and accumulates counts of the five types of burst errors."[102] The Media Diagnosis screen once again requests the Volume and Side IDs. The results of the media diagnosis will appear in the common message area. Both sides of each optical disk should have this procedure performed on them prior to the disk's use for optical storage.

The Mount Volume (MV) command mounts a disk in the jukebox or one of the two stand alone drives. The Mount Volume screen requires the operator to provide the desired Volume, Side, and Unit IDs. Upon entry of this information the disk is mounted in the appropriate drive with the indicated side facing up. While stored in the jukebox information on both sides of any mounted disk is viewable. However, in stand alone drives only data from the side indicated on the Mount Volume screen may be retrieved. In order to view material on the other side of a disk mounted in a stand alone drive, the disk must be dismounted and then remounted with the ID of the desired side indicated in the Mount Volume request. The

---

[100] *Unisys Optical Digital Image Storage System (ODISS), Volume XI - ODISS Operations Manual*, page 5-10.

[101] *ibid*

[102] *ibid*

results of mount volume requests are displayed in the common message area. The MV command is among the most commonly used features of the Archive Control Terminal.

The Volume Status (VS) request returns the Archive Volume Status screen showing the drive location of all mounted volume sides. For volumes mounted in the jukebox, shelf location is also given. Volumes stored in the present jukebox must be mounted in unit j1; however, they will be listed as located in unit j0 on the Archive Volume Status screen.

Functions other than those provided by the IMS commands may be performed at the Archive Control Terminal by pressing the DELETE key on the keyboard. A prompt will then appear. If a problem occurs with one of the IMS screens, typing IMS after the prompt will bring the Archive Control screen back up.

The most frequently performed non-IMS functions are core system reboots and resets. These procedures are performed in the case of a system-wide hang, crash, or erratic behavior pattern.

The core system reboot requires powering off the core system, powering on the core system, and resetting UNIX. When UNIX is reset a Heurikon ">" prompt appears on the screen of the Archive Control Terminal, and the remaining reboot procedures should then be followed. It is not necessary to bring up a prompt by pressing the DELETE key in order to reboot the core. The System Manager system may be rebooted from the System Manager Terminal, but this function is not frequently performed.

Before doing a full core reboot, it is first preferable to attempt a VRTX reset. This is a less extensive, and thus less time-consuming, procedure than the core reboot. The VRTX reset may be accomplished by resetting the core VRTX and then reinitializing ODISS from the Archive Control Terminal. A prompt must be brought up by pressing DELETE, so that the reset commands may be entered at the Archive Control Terminal.

Disk diagnosis which is performed on each disk side, once the side is completely full. This program enumerates files on the disk side by the number of the file on the disk, sector, and File Control Number. Data on next sector available for archiving, full sectors in data area, deleted sectors in data area, and full sectors in directory area are summarized at the end of the test results. The diagnosis results are displayed on the CSE/ARS terminal under its ARS function.

Disks on which diagnosis is to be performed should be mounted in one of the stand alone drives with the side to be tested placed face up. The disk diagnosis program overrides the regular IMS program; for this reason it is preferable to leave a disk diagnosis running overnight rather than to run it during normal working hours.

B.12.3 System Manager Duties

ODISS had a system manager computer function, and a personnel system manager in a supervisory position. The position of system manager requires performance of a range of activities, mostly, but not entirely related to the operation of the three terminals - System Manager, CSE/ARS, and Archive Control - located at the System Manager station.

Among the principal duties of the system manager is the responsibility for making additions, deletions, inquiries, and modifications to various files on the System Manager database as

such actions are needed. The system manager may also be required to delete files from magnetic disk at the CSE terminal or from optical disk at the Archive Control terminal.

The system manager periodically consults the various terminals to determine the general flow of material through the system. Consultation of the BLOCK and Blocks Ready to Archive records on the System Manager terminal and the Volume Status display on the CSE terminal are especially helpful in performing this function. Observation of the physical disposition of blocks within the room, principally on the shelves reserved for quality control and rescan boxes, may also be useful. Information gleaned from such studies suggests that reapportionment of employees among the various workstations is in order.

Initiation of standard programs and production of routine reports are other standard duties of the system manager. Typical programs which the system manager must run include the "FCNLIST," "FCN2LIST," "NUMCHECK," and "NAMELIST" programs on the System Manager Terminal and the disk diagnosis on the Archive Control Terminal. The most common computer reports are the daily, weekly, and quarterly reports on the various workstation elements. The system manager may from time to time be requested, or deem it useful, to compile manually written reports from various ODISS sources.

The system manager also maintains various manual records. The principal among these is the Archive Block Status log, a log listing the status of all blocks between completion at the rescan station and final archival to optical disk. This log prevents confusion, especially in the absence of the regular system manager, about which of the final steps prior to its archive have been undertaken for each block. Additional manually maintained records include accounts of file and block deletions from magnetic disk, file deletions from optical disk, namelist-type data for partial blocks entered at the rescan station, and test results from disk diagnoses.

The system manager is also responsible for the availability of certain forms relating to normal ODISS operations, such as the Quality Control Log, a record of blocks quality controlled by each operator; and, on occasion, a block action sheet, a listing of files within a block requiring specific rescan actions.

The system manager has charge of the final preparatory acts prior to an archive. These include performing or overseeing the running of the pre-archive and/or "NUMCHECK" programs; running of the "FCN2LIST"; comparison of folder labels with the "NAMELIST"; and if necessary correction of the CMSR database. When required, the system manager may return a block designated for archive to any workstation element in the system.

The most important responsibility of the system manager, however, is initiation of the final archival to optical disk. The system manager is also responsible for the creation of duplicate copies of optical disks. Further, the system manager has general overall responsibility for the mounting and dismounting of optical disks.

Another cardinal function of the system manager is production of the database backup. This program should be initiated daily from the System Manager terminal.

The system manager also provides assistance to NNTH supervisory personnel, NSZ staff, and ODISS contractors. The system manager provides them, orally or in writing, with reports or data for reports; gives them timely notification of any problems or potential problems, either machine or staff related; and supplies general advice and commentary on ODISS.

APPENDIX C

COMPILED SYSTEM PERFORMANCE DATA

APPENDIX C. COMPILED SYSTEM PERFORMANCE DATA

C.1 Tennessee CMSR File Sample

System performance data was captured during the test of converting Confederate Tennessee compiled military service records (CMSR). The processing of CMSR records began shortly after the installation of ODISS in July, 1988. The conversion of the Tennessee cavalry files by NN's ODISS operations staff began August 8, 1988, and concluded May 26, 1989. This time span included 201 work days. The last few files were completed by a single remaining NN staff member in late May and early June, 1989.

The original plans for the test were to convert the entire set of Tennessee CMSR. The test was concluded after all the Tennessee cavalry regiments had been written to optical disks. In early May 1989, it was concluded that it was unnecessary to continue with the conversion of the infantry and artillery unit files since the cavalry files had provided a sufficient quantity of data to determine the aspects and implications of CMSR conversion.

C.1.1 Quantity Converted

The Tennessee calvary consists of 76 regiments with 798 companies. Six of these regiments do not have any subordinate companies. The records of the 76 regiments were written to five 12-inch two-sided optical disks and a quarter of one side of a sixth disk.

The conversion system with its automated counting feature gave NARA its first precise figures for the number of CMSR records. The Tennessee cavalry CMSR records were found to consist of almost 54,000 files. When converted to optical disk, these files proved to contain more than 220,000 images. The total numbers of files and images in the Tennessee Confederate cavalry on optical disk as of June 8, 1989 were:

Quantity of CMSR Records Converted

Number of Files	Number of Images
53,783	220,713

Table C-1

C.1.2 File Size

The Tennessee cavalry records average 4.1 images per file. This file size is much smaller than was anticipated. The 1984 report on pension, bounty-land, and compiled military service records utilized an earlier GSA/NARS survey of the records that estimated the average size of CMSR files as 15 pages. This proved widely erroneous for the Tennessee calvary. The number of files with 15 images is only 362. Another unexpected finding was that there are 9,975 single-page files consisting only of a cross reference card; these comprise 18.5% of the cavalry files. The numbers of Tennessee cavalry files in different size ranges are shown in Table C-2.

Ranges of CMSR File Sizes	
Size Range of File	Number of Files
1 - 5 images	42,923
6 - 10 images	6,668
11 - 20 images	3,668
21 - 30 images	377
> 30 images	160

Table C-2

C.2 Conversion Statistics

ODISS includes automated data collection capability for the major actions at each of the conversion functions and workstations. The data can be processed to generate management reports for different time periods defined by the user. These statistics gathered by the system are the basis for quantitative information about the work done during the CMSR conversion.

The data for overall conversion activity does not exactly match the number of files and images actually written to optical disks. From July into November, 1988, the system had many problems that often required work to be redone at one or more functional stages before files were ready to write to optical disk. The system's automatic data collection function cannot distinguish or delete the numbers for the many stops and restarts during the shakedown months. Although the Unisys staff sometimes cleaned out the database, including the conversion statistics, after they fixed problems, this was not always done.

Moreover, in the early months, there were some problems with improper sign-off procedures by the operators that at times caused the system to fail to record a session's statistics. In the fall of 1988, the low speed scanner had a problem that prevented it for collecting production statistics. These various problems mean the totals at each step in the conversion system do not exactly match each other. Because much work had to be redone for certain functions, especially in the early months of the project, the total numbers for files and images captured in the management reports as processed by the conversion system are larger than the numbers for files and images finally written to optical disk.

Another complication to statistical accuracy stems from the fact that while the full conversion staff ceased work after May 26, 1989, more work was done during the ensuing month by one person who was left from that staff to complete files still in the pipeline and to redo a few problem files. The figures for the work done after May 26 are small enough not to have a statistically significant impact on the conversion totals. The figures used here are limited to the data from the work performed by the full NN operations staff between August 8, 1988 and May 26, 1989.

C.2.1 High Speed Scanner Totals

The system measured the number of images scanned and the number of files processed at the high speed scanner. The totals for the period between August 8, 1988, and May 26, 1989 are shown in Table C-3.

High Speed Scanner Production	
Files Processed	Images Scanned
54,394	232,846

Table C-3

C.2.2 Indexing Totals

The system measured the number of files indexed. The initial workflow plan called for the use of two workstations for indexing and the assignment of other workstations to the function when backlogs developed. Backlogs did occur at various times, and indexing was done at more than two workstations as needed. The number of files indexed between August 8,1988 and May 26, 1989 was:

Indexing Production	
Files Indexed:	54,746

Table C-4

C.2.3 Quality Control Totals

The workflow plan assigned two workstations to quality control and provided for assigning other stations to the function when backlogs occurred. More than two stations were assigned to quality control on various occasions.

The operators compared all the documents in all the files with their images on the workstations' display screens. When the images were hard to read, the operators put the paper document in a colored folder and used an electronic tag on the digital image to mark it as rejected and needing rescanning at the low speed scanner. If the operator found a paper document for which no image existed, the document was put into a colored folder and an electronic "not scanned" tag was placed in the digital file to indicate the need and location for inserting an image using the low speed scanner.

The system counted the actions taken at quality control. There are statistics for the number of the files approved as having no image quality problems and the files rejected for image quality problems, as well as the total images reviewed and the images rejected for poor quality. The system also counted the number of pages in the paper file that the operators marked as not scanned at the high speed scanner. The figures for August 8, 1988 through May 26, 1989 are shown in Table C-5.

Quality Control Production	
Files Approved	50,152
Files Rejected	8,350
Images Reviewed	256,948
Images Rejected	15,660
Images Not Scanned	1,336

Table C-5

C.2.4 Low Speed Scanner Totals

At the low speed scanner two kinds of work were done, original entry and rescanning. Files that could not be scanned successfully at the high speed scanner were entered at the low speed scanner in its "original entry mode." The system counted work done in the original entry mode as files processed and images scanned.

The bulk of the work at the low speed scanner was the rescanning of poor images marked at quality control. The system counted this work as files redone and images rescanned.

A problem at the low speed scanner from October 28 through November 16, 1988, prevented the system from collecting conversion data for this function. Consequently, the available low speed scanner numbers are lower than its actual production. The low speed scanner's figures for August 8, 1988 through May 26, 1989 were:

Low Speed Scanner Production	
New Files Processed	935
New Images Scanned	4,266
Files Redone	6,326
Images Rescanned	12,765

Table C-6

C.3 Analysis of the Conversion Statistics

Despite their variations, the statistics for converting the Tennessee cavalry files offer good clues to several aspects of CMSR conversion in general.

C.3.1 Optical Disk Storage Capacity

Industry claims about the high storage capacity of optical disks proved valid for the CMSR records. ODISS uses 12-inch, two-sided SONY disks, and they accepted an average of more than 20,000 CMSR images per side or over 40,000 images per disk. For the five disks that

were completely filled during the conversion, the numbers of CMSR images per disk are shown in Table C-7.

Images Written to Optical Disk

Disk Number	Images on Disk
1	41,422
3	38,273
5	43,130
7	44,097
9	47,609

Table C-7

The average for these five disks was 42,906 images per disk.

C.3.2 High Speed Scanning

Before the installation of ODISS, there was some fear expressed that many, and perhaps even most, CMSR documents would be too fragile to send through the high speed scanner. This proved to be an unnecessary worry since the high speed scanner processed almost all of the CMSR files and documents without harm.

One indication of this is the small number of files that had to be processed at the low speed scanner's original entry mode. While 53,783 files were written to the optical disk, only 935 files had to enter the system through the low speed original entry mode. Therefore, only about 1.7% of the CMSR files could not be processed at the high speed scanner.

Another question concerned the thoroughness of the high speed scanner operators in processing all the documents in the files. To measure the rate of documents missed at the high speed scanner, the quality control operators were given the capability to add electronic marks to the files for pages not scanned. The total "not scanned" figure was only 1336. As a percentage of the 256,948 images reviewed, this amounts to only 0.5%. In other words, pages being missed by the high speed scanner operators was a very minor problem.

C.3.3 Image Quality Rejection Rate

One goal of the ODISS test was to learn how many images would be judged unacceptable at quality control and require rescanning for better image quality at the low speed scanner. The operators were instructed to mark images for rescan if the images were not legible. The statistics collected at quality control showed 15,660 images were marked for rescan out of 256,948 images that were reviewed. This is a rejection rate of 6%. The remaining 94% of the CMSR pages sent through the high speed scanner did not require any further work to make their images more legible.

C.3.4 Average Daily Production Rates

The conversion period for Tennessee CMSR files from August 8, 1988 through May 26, 1989 consisted of 201 available work days. However, daily management reports generated from the automatic data collection by ODISS do not show exactly 201 days of activity at all of the major input functions. For example, the reports show only 189 days of work at the high speed scanner, and the reported production for one of these days was 0. For indexing, there are management reports for all 201 days, but three of the reports showed 0. Another example is the low speed scanner where the reports were inoperative from late October through mid November, 1988.

Daily production can be calculated for either the full 201 available work days or for the number of days worked at each input function as indicated by the management reports. Whatever basis is chosen for the calculation, the daily production of ODISS was much lower than was expected. The production rates of the ODISS system with its present software flaw which retard throughput would not make feasible the efficient conversion of all the CMSR records.

C.3.4.1 Rates Based on the Full Available Work Period

Using the full 201 available days to calculate production rates has the advantage of providing a standard time period to measure all the input functions on the same scale. This approach ignores some of the interpretive problems stemming from variations in the data, but it does tell what was accomplished on average throughout the whole period.

The daily average production rates for the major conversion operations based on 201 work days are shown in Table C-8.

C.3.4.2 Rates Based on the Days with Management Reports

As indicated in the previous section, the daily management reports for some functions did not cover all 201 work days. For example, there were a number of times when the high speed scanner was not run because of backlogs at the other operations in the system. Moreover, for some days, the management reports showed 0; this occurred when a staff member logged onto the system for that function but did no work before logging off. If all the days when no work was performed are subtracted from 201, the result is the number of days of active operations. Then the production rates of each input function as a daily average for the time of active work can be calculated.

The figures for average daily production for the days of active operations as indicated by the management reports are shown in Table C-9.

Average Daily Production Rates - All Work Days	
High Speed Scanner:	
Files processed	271
Images scanned	1158
Indexing:	
Files indexed	272
Quality control:	
Files approved	250
Files rejected	42
Images reviewed	1278
Images rejected	78
Images found "not scanned"	7
Low speed scanner - rescan mode	
Files redone	31
Images rescanned	64
Low speed scanner - original entry mode:	
Files processed	5
Images scanned	21

Table C-8

Average Daily Production Rates - Active Days	
High Speed Scanner (188 active days):	
Files processed	289
Images scanned	1239
Indexing (198 active days):	
Files indexed	276
Quality control (199 active days):	
Files approved	252
Files rejected	42
Images reviewed	1291
Images rejected	79
Images found "not scanned"	7
Low speed scanner - rescan mode (185 active days):	
Files redone	34
Images rescanned	69
Low speed scanner - original entry mode (103 active days):	
Files processed	9
Images scanned	41

Table C-9

APPENDIX D

COST ANALYSIS

APPENDIX D. COST ANALYSIS

D.1 Cost Analysis Methodology

This appendix presents a cost comparison of various records conversion alternatives with a baseline consisting of maintaining a set of original paper records and providing reference services directly from those original records.[103] In order to standardize this analysis, a model consisting of a generic, hypothetical application was used. The six approaches compared are:

- Existing Paper Storage and Retrieval Operations (the baseline)
- Simple Microform Conversion and Retrieval Operations
- Upgraded Microform Conversion and Retrieval System
- Upgraded Microform Using Service Bureau Conversion
- Digital Image/Optical Disk System
- Digital Image System Using Service-Bureau Conversion

The first two alternatives make use of existing paper, microfilm conversion, and reference capabilities, facilities, and operations. The third and fourth alternatives involve installation of computer-assisted and retrieval technology. The remaining two alternatives involve acquisition and use of digital image and optical disk storage systems.

D.1.1 Generic Application Description

In order to obtain comparative results from the various system alternatives, analytical parameters were standardized within the application model. This hypothetical, generic sample contains 400,000 files consisting of eight million documents, one half of which are two-sided. This results in a total of twelve million images. Seventy-five percent are office-type documents, easily handled by a high speed paper transport. The balance are oversized, extremely fragile, or in very poor condition requiring special handling during conversion. The average file size is twenty pages (thirty images using the 50% two-sided factor).

The generic model's reference activities include a "custodial unit" with five staff members capable of processing two hundred mail requests daily for manual paper and microfilm systems. The microfilm and digital image systems can process the daily reference load using three staff members and one staff member respectively.[104] An average of ten walk-in

[103] The ODISS project involved a retrospective conversion of paper records to an alternate medium (optical disk.) For the sake of consistency, the alternatives used in the cost comparison all involve a conversion of paper documents to an alternate medium. The analysis does not consider other possibilities such as creating a computer-based index for use with the original paper records.

[104] Reference staff requirements are based upon NARA performance standards for paper and microfilm reference operations. Requirements for optical disk systems are based upon statistical data from the ODISS project.

requests are received and 4,000 hardcopies are produced daily in response to mail-in requests.[105] All reference is supported from a single site at which there are both staff and public retrieval stations. The staff and public stations are located in the same building although not necessarily in the same workspaces.

Conversion must be completed within a three-year time period for each of the five approaches that require a conversion. System performance figures are based upon seven-hour work days, five days each week, fifty weeks a year, over a three year period, for a total of 5,250 hours. Costs for each system alternative have been calculated over a ten-year period. Table D-1 summarizes the general attributes of the model used for all six system alternatives.

Generic Study Model

Application Universe

- 8 million pages of which 50% are double-sided
- 12 million total images
- 36 month in-house conversion time allowance

Document and File Characteristics

- 6 million pages (75%) are high speed transportable
- 2 million pages (25%) require special handling
- 20 pages (30 images) per file
- 50 kilobytes average compressed image data file size (for digital image systems)
- 10 index fields are manually key entered (for computer-indexed systems)

Retrieval Characteristics

- 200 mail requests per day per unit
- 10 walk-in requests daily per unit
- 4,000 hardcopy pages per day per unit

Table D-1

[105] The 4,000-page reproduction figure is derived from NARA experience that only about 62% of the reference requests actually result in a successful search and a subsequent demand for copies of the records.

D.1.2 Data Acquisition

As for any cost analysis, background data was drawn from available sources. The ODISS project team gained considerable digital image technology experience from ODISS and other outside studies. Design configurations of the various options were equalized as much as possible using the most reasonable and conservative figures available. Technical specifications and cost figures for existing equipment and systems were obtained from related studies and reports. The computer-assisted microform and digital imaging systems hardware and integration costs were based upon vendors' quotes and price lists for equipment meeting generic technical specifications. Microfilm supply and equipment costs were taken from GSA schedules. In order to standardize the cost data, the calculations utilize constant dollars and workload figures. The cumulative cost comparison table (Table D-8 on page 320) presents system costs in terms of "Net Present Value."[106]

D.1.3 Assumptions and Constraints

As with most cost models, this one is constructed around a set of facts, assumptions, and constraints. For example, the model presented in this appendix is constructed around a hypothetical archival collection, the attributes of which were outlined in section D.1.1. Arbitrary dimensions were also given to the time allowed for conversions (three years) and to the size of the retrieval workload. Changes to any of the input parameters in the model would necessarily alter the model. Therefore, the model should be viewed, considered, and interpreted in light of these assumptions and constraints.

Assumptions and constraints which apply to features or factors used in more than one of the comparison alternatives are listed below. Ones that apply only to a particular alternative are listed in the section where that alternative is presented.

- The model uses a paper reference system as a baseline. It presumes that the typical archival institution can already support reference from original paper records. It also presumes that the typical institution has basic microfilming facilities. Therefore, it is presumed that there is no need for initial outlays for equipment or facilities to support the first two alternatives outlined in sections D.2 and D.3. In these and all options, the costs identified are incremental costs in arriving at the stated level of capability from a base level provided by those equipments, systems, or conditions which are presumed to pre-exist.

- No authoritative source could be found for labor and archival storage costs on the basis of a nationwide average. Therefore, this model uses labor and storage costs identified for the National Archives in Washington, DC. It is acknowledged that Washington, DC may be considered a high-cost area and that the figures used in the table may reflect this. For example, reference staff salaries are shown as $19,200 per year. Yet an informal survey of several state archives indicates that this wage figure may be somewhat low. Experienced reference staff, who typically are the most familiar with the records holdings, often receive higher wage compensation. Annual

[106] *Net Present Value* is an accounting method used in determining payback periods which takes into account the time value of money. In these types of determinations, OMB Circular A-94 prescribes the use of a 10% discount rate based upon "... an estimate of the average rate of return on private investment before taxes and after inflation." For further discussion, refer to Section D.8.

salaries of professional research staff may exceed the mid-thirty thousand dollar range, excluding fringe benefits. Less experienced staff earn reduced salary amounts. Salary costs are influenced by geographic location, job description, and employee's service record. Fringe benefits may also vary widely. Wage rates for microfilm and digital image equipment operators are listed as $16,905 per year. This is based upon the federal government pay schedules for GS-04 employees. The survey revealed that it is possible for experienced equipment operators to exceed $20,000 annually, excluding fringe benefits. Equipment operators with limited experience receive correspondingly lower annual salaries, while production line supervisors may earn significantly more.

Document storage costs are also affected by unique regional conditions. Storage figures used in the tables are $2.50 per cubic foot, resulting in total yearly costs of $8,000 for the eight million documents in the model. Storage costs should be tailored to local rent or building maintenance conditions. The costs for storing documents in older, established archives may differ from equivalent space in modern office complexes. The tables clearly demonstrate the cost savings available by storing records in less convenient remote facilities. The costs of using the underground storage facilities at Boyers, Pennsylvania were selected as representative.

In summation, any reader wishing to use or interpret the model should understand that substitution of his or her local labor and storage costs into the model may alter its results.

- Most equipment identified in the cost analysis, especially those with high-technology, electronic, or computer-based components, are presumed to have a useful life of seven years. Consequently, the cost charts show the replacement of those equipments and related software and services in year seven. Considering the advances in technology that generally take place in that timeframe, it is expected that higher levels of capability will have become available for less money at the time that the systems are to be replaced. Therefore, for high-technology equipment, a reduction factor of 50% has been applied to the replacement systems. No reduction factor has been applied to equipment from mature technologies (e.g., microfilm readers) or to software or services which have labor-intensive components.

- Annual equipment and software maintenance fees have been estimated at ten percent of the original cost of the equipment and software. This follows an accepted industry rule-of-thumb.

- In the five alternatives involving a conversion of paper records to an alternate medium, certain activities and their associated costs are phased in or phased out during the conversion period of three years as more and more records have been taken through the conversion process. Typical examples would be the phasing out of reference operations using the original paper records as those records are converted and retired; and the phasing in of reference operations using the records in their converted form (i.e., microfilm or digital image). It is presumed that these costs would decrease or increase linearly over the period of the conversion. Therefore, in these cases, the cost analysis charts use mid-year costs to average the costs for the entire year.

- The cost tables reflect the assumption (based upon NARA performance standards and production statistics) that the alternatives involving automated or partially automated systems would require fewer personnel to handle the reference workload posed by the model holdings, whereas manual systems require more reference personnel to handle the same workload using labor intensive search processes.

- The cost tables for the alternatives involving new automated systems include a one-time cost of $5,000 to acquire user training materials and services, an expense not needed with the existing manual operations.

- Hardcopy print and microfilm duplication costs in conjunction with reference services are not used in the cost tables for any of the six options. It is presumed that costs for the labor and materials to produce copies for researchers would be fully recovered through user fees.

D.2 Existing Paper Records System

This alternative is used as the baseline against which to compare the five remaining alternative systems. In it, the majority of the existing staff reference services is accomplished with paper records. This section calculates costs of storing, preserving, and performing reference services using the model records sample with reference procedures similar to those used at NARA. The calculations are based upon cost data provided by the Office of the National Archives, augmented with information contained in fee schedules provided by the Office of Management and Administration.

D.2.1 Description

This alternative presumes the usage of current facilities and methodologies to support records maintenance and reference services using original paper records. It involves costs for storage, preservation, and staff reference. Five reference staff would service the mail-in requests using the original paper records with photocopies produced and mailed to the researchers. The walk-in research requests would be serviced from the original paper records. There are no conversion requirements or costs associated with this alternative.

D.2.2 Derivation of Costs

This option has three primary cost categories: document storage, document preservation, and reference staff.

- <u>Storage</u>: Using NARA's standard storage estimate of 2,500 documents per cubic foot, eight million pages requires 3,200 cubic feet of shelf space. With storage costs estimated at $2.50 per cubic foot,[107] the resulting annual storage cost for the model collection would be $8,000.

---

[107] The cost figure of $2.50 per cubic foot used in this cost analysis was obtained from the Office of the National Archives as the average of storage costs at the eleven regional archives. (In the Washington, DC area, the cost of leased storage space at NARA's Pickett Street facility is $13.80 per square foot. This converts to $2.76 for five-high shelving or $2.30 per cubic foot for six-high shelving.)

- Preservation: Document preservation consists of holdings maintenance and document conservation activities. For the model collection, maintenance costs are \$19.89 per cubic foot and conservation lab treatment costs \$2.32 per document.[108] This equates to \$73,539 for the eight million documents over three years.[109]

- Reference Staff: Referencing this eight million document collection under the generic model's structure requires five full-time staff at \$19,200 each. The reference staff perform all required functions including finding aid/index lookups, records retrieval, photocopying, and records refiling. Two hundred mail requests and ten walk-in requests are handled daily equating to forty mail requests and two walk-in requests per staff member. In addition, four thousand pages are photocopied each day to service researcher requests.

Ten-year cost figures for supporting this option are presented in Table D-2.

D.2.3 Advantages and Disadvantages

Advantages

- No conversion costs are incurred, paper remains in existing form.
- Researchers and staff are familiar with system; no learning of new systems or procedures is required.
- No recurring costs for computer workstation repairs or upgrades.
- Paper records storage systems can be rudimentary shelves with minimal locator guides.

Disadvantages

- Prime storage space is occupied with paper holdings.
- Reference services require labor intensive staff activities.
- Misfiled paper records with limited indexes are essentially lost.
- Deterioration to archival documents caused by general public handling.
- Records refiling requires additional staff time.

---

[108] For this model, the quantity of documents requiring conservation treatment was 4,263.

[109] Holdings maintenance and conservation laboratory costs were obtained from the Office of the National Archives FY 1989 preservation program cost tables.

Paper System - Cost Breakdown
(Actual Costs)

COMPONENT	YEAR 1	2	3	4	5	6	7	8	9	10	TOTAL
STORAGE	$8,000	$8,000	$8,000	$8,000	$8,000	$8,000	$8,000	$8,000	$8,000	$8,000	$80,000
PRESERVATION	$24,513	$24,513	$24,513								$73,539
REF. STAFF	$96,000	$96,000	$96,000	$96,000	$96,000	$96,000	$96,000	$96,000	$96,000	$96,000	$960,000
TOTALS	$128,513	$128,513	$128,513	$104,000	$104,000	$104,000	$104,000	$104,000	$104,000	$104,000	$1,113,539

Table D-2

D.3 Manual Microfilm System

This option entails the conversion of twelve million images using existing microfilming equipment, personnel, and processes. This is a totally manual storage and retrieval system, with no computer-assisted capabilities. This option presumes that basic microfilming equipment already exists on-site and was procured sufficiently long ago that it is fully depreciated, resulting in no start-up costs for conversion equipment. In order to produce 4,000 hardcopies each day, four new microfilm reader/printers were considered essential. Supplies, conversion and reference staff, and yearly recurring storage and retrieval costs are the major overall expenses for this option.

D.3.1 Description

Three trained staff will prepare the documents[110] over a three-year period conversion period. The prepared documents will be converted to 16mm microfilm. Up to 2,500 documents will be captured on each 100-foot roll of microfilm using hand-fed planetary cameras. Using a daily production rate of 3,331 images per day for fully prepared flat work, a three-year conversion of twelve million images will require five planetary cameras. The exposed microfilms will be developed using a table-top processor and checked for technical quality. Accepted master microfilm will be copied onto direct duplicate silver print materials. Two 16mm microfilm duplicates will be prepared for staff and public reference usage. Since automated indexing is not involved, no costs are incurred for staff time or equipment to key enter image locations.

Two duplicate sets of microfilms are accessible in the staff and public user areas for information retrievals. Reference staff or public walk-in users would consult a manual index, and manually load and advance the film to locate the desired images. Hardcopy prints would be produced, and the film rolls would be returned to the storage cabinet upon completion of the search session. The original documents and the 16mm master microfilms would be transferred to low-cost, environmentally stable storage.[111]

D.3.2 Derivation of Costs

Except for the four new readers/printers, there are no expenditures for microfilming hardware as this system uses cameras, processing, and duplication equipment which are assumed to exist already at most if not all archival institutions.[112] Microfilm and processing chemistry supply costs are derived from current GSA supply schedules. The conversion staff consists of five camera operators, supervised by one task leader.

---

[110] Document preparation involves document repair, paper flattening, staple removal, insertion in clear polyester sleeves, and other activities to get documents ready for conversion.

[111] NARA's film masters are maintained in underground storage at Boyers, Pennsylvania.

[112] Costs will be higher than those shown in Table D-3 if the needed equipment is not available and must be purchased.

Explanations of nomenclature and derivation of certain costs used in Table D-3 follow:

- Camera Film (16mm): 4,800 rolls (100 ft.) of microfilm are needed to record 12 million images; costs spread over three years of conversion.
- Processing Chemicals: Microfilm processing solutions; volumes based on daily use and replenishment; costs spread over three years; costs in year seven are for six month supply of chemicals needed to produce two replacement sets of positive prints.
- Dir. Dup. Neg: Duplicate negatives used to produce succeeding microfilm positive prints.
- Pos. Prnt: Duplicate microfilms for staff and public reference; because of wear and tear, would be replaced at year seven.
- Document Prep Staff: Eight million documents require three GS-04 technicians ($16,905/yr) for three years to prepare the records for conversion. As noted earlier, federal wage scales were used where more widely representative figures were not available.
- Doc Prep Mat & Srvcs: $10,000 each for materials needed in document preparation and conservation laboratory services to repair damaged documents discovered during document preparation; costs spread over three years.
- Cnvrsn Staff: Six staff members are needed: five to operate the planetary cameras and one to operate the film processors and duplicators. A supervisor is also required.
- Paper Prsrvtn Costs: Paper holdings maintenance and conservation laboratory costs for three years.
- Paper Storage: Converted documents would be transferred to lower-cost underground storage as conversion is performed; costs phased in during first three years.
- Film Reels/Cartons: 14,400 boxes and plastic reels to hold the duplicate negative and staff/public microfilms; costs split evenly over three years of conversion.
- Cur. Paper Sys. Cost: Three years of linearly declining costs are included for paper records storage and reference system operation; mid-year costs are shown.
- Film Stor. Cabinets: Microfilm storage cabinets for the staff and public microfilms.
- Film Reader Prntrs: New reader/printers for staff and public users, replaced after seven years.
- Master Film Storage: 4,800 rolls of camera master microfilms stored underground under environmentally stable conditions.
- Reference Staff: The manual microfilm system requires three full time staff to handle user requests.

- Dup. Neg. Storage: 4,800 rolls of direct duplicate negative microfilms require on-site storage; first three years show mid-year costs as films are produced during the conversion.

- Equip. Maint.: Costs for yearly equipment maintenance based upon 10% of original equipment costs.

- Film Prsrvtn/Inspctn: Inspection program for camera master films to ensure archival quality and longevity; first three years show mid-year costs as films are produced during the conversion.

Ten-year cost figures for this option are presented in Table D-3.

D.3.3 Advantages and Disadvantages

Advantages

- Microfilming protects the documents by eliminating records handling by researchers.
- Archivist staff is released from manual pulling and refiling of records.
- Since mostly existing equipment is used, little additional hardware expense is involved.
- Use of manual microfilm facilitates the production of copies that can be used at remote sites without the need for special equipment (other than conventional microfilm readers.)

Disadvantages

- Manual look-ups are needed due to absence of computerized, automated indexes.
- Existing conversion and retrieval equipment is old, some of it obsolete.
- Limited staff available for massive in-house microfilm conversions.

Manual Microfilm System - Cost Breakdown
(Actual Costs)

COMPONENT	QTY	EACH	YEAR 1	2	3	4	5	6	7	8	9	10	TOTAL
CAMERA FILM (16MM) [16MM X 100 FT]	4,800	$4.68	$7,488	$7,488	$7,488								$22,464
PROCESSING CHEMICALS (BY LOT)	1		$6,089	$6,089	$6,089								$18,267
DIR.DUP.NEG. [2500ft ROLLS]	200	$60.22	$4,015	$4,015	$4,015								$12,044
POS. PRNT. (16MM) [2500ft ROLLS]	400	$49.70	$6,627	$6,627	$6,627	$20	$20	$20	$20	$20	$20	$20	$20,019
DOCUMENT PREP STAFF	3	$16,905	$50,715	$50,715	$50,715								$152,145
DOC PREP MAT & SRVCS			$6,667	$6,667	$6,667								$20,000
CNVRSN STAFF [GS04]	6	$16,905	$101,430	$101,430	$101,430								$304,290
[GS06]	1	$21,081	$21,081	$21,081	$21,081								$63,243
PAPER STORAGE			$873	$2,620	$4,367	$5,240	$5,240	$5,240	$5,240	$5,240	$5,240	$5,240	$44,540
FILM REELS/CARTONS	14,400	$0.50	$2,400	$2,400	$2,400								$7,200
CUR.PAPER SYS.COST			$86,667	$52,000	$17,333								$156,000
FILM STOR.CABINETS	8	$1,313	$3,501	$3,501	$3,501								$10,504
FILM READER/PRNTRS	4	$12,988	$51,952						$51,952				$103,904
MASTER FILM STORAGE	4800	$0.16	$128	$384	$640	$768	$768	$768	$768	$768	$768	$768	$6,528
REFERENCE - STAFF	3	$19,200	$9,600	$28,800	$48,000	$57,600	$57,600	$57,600	$57,600	$57,600	$57,600	$57,600	$489,600
DUP.NEG.STORAGE	4800	$0.18	$144	$432	$720	$864	$864	$864	$864	$864	$864	$864	$7,344
EQUIP.MAINT.(BY LOT)	1		$5,195	$5,195	$5,195	$5,195	$5,195	$5,195	$5,195	$5,195	$5,195	$5,195	$51,952
FILM PRSRVTN/INSPCTN	4800	$0.16	$128	$384	$640	$768	$768	$768	$768	$768	$768	$768	$6,528
TOTALS			$364,700	$299,828	$286,908	$70,455	$70,455	$70,455	$122,407	$70,455	$70,455	$70,455	$1,496,572

Table D-3

D.4 Upgraded (CAR) Microfilm System

This section estimates the costs and equipment needed to acquire a totally new microfilming system utilizing high speed microfilming for document conversion and computer-assisted retrieval (CAR) technology for reference. All production steps to microfilm the twelve million images would be performed in-house by staff.

D.4.1 Description

The record holdings would be prepared for filming by trained staff. This option involves new equipment obtained especially for this project. The system uses high speed and low speed cameras, thin base films for greater image compaction, indexing and quality control stations, and computer-assisted image retrieval (CAR) devices. The filmed images would be indexed and checked for acceptability and completeness in a dual-purpose workstation. Silver halide duplicates would be produced from the camera master films. The staff and public workstations would each have a roll film library. Hardcopy prints would be created on user demand at a theoretical rate of 4,000 per day. Keyboard controllers would be used to search and retrieve the desired frames on the user workstations, supported by a computerized database accessible from each workstation.

Microfilm Cameras

A high speed microfilm camera system would capture 75 percent of the documents that are physically sturdy enough to withstand the rigors of automated transport. This camera would use 16mm microfilm with a polyester base 2.5 mils thick. Images would be captured at 24X reduction, yielding approximately 5,260 images per 215-foot roll. A minimum camera production rate of 1,715 images per hour would be needed, and approximately 2,280 film rolls would be required to capture 12 million images.

A low speed camera would capture the fragile documents and also refilm any defective images discovered during quality control inspections. This tabletop microfilmer, requiring hand-placement of fragile documents, would capture two million pages within the three-year production period.

Film Processing

The exposed microfilms would require chemical development to obtain visible, permanent images. Using equipment assumed to preexist at most any archival facility that has a basic microfilming program, a technician would process all films, create silver duplicates, and perform technical image quality inspections. Rejected images would be refilmed.

Computer Database Index Creation and Image Inspection

A microfilm CAR system automatically searches the database and identifies the desired frames. The index station would use a computer-linked video display and keyboard, and an automated blip-counting[113] film viewer. The film viewer would stop at each frame so that the index data could be extracted and key-entered into the computer. Two workstations could

[113] Blip encoding of microfilms involves small machine-readable marks under images, used by the automated retrieval units to locate the desired frames.

meet the time requirements of the 400,000 files to be indexed and quality checked over the three-year conversion period. The indexing requirements are for key entry of ten information fields. Images requiring refilming would be noted so that the original documents could be refilmed and spliced into the rolls as needed.

Duplicate Microfilm Production

A [preexistent] silver printer would be used as a high-speed roll-to-roll device, accepting large-capacity 2,500-foot rolls of duplicate film supplies. The developed prints would be inspected and wound onto reels using an editor/loader. User copies would be loaded in ANSI/AIIM approved magazines for CAR retrieval. Two positive prints would be required, one print for the staff workstations, and the second print for the public user workstations. Camera master films would be stored in archival quality containers. Following filming and duplication, documents would be maintained off-site.

Staff and Public User Reference

Staff and public would be provided with a complete set of microfilms in non-motorized carousel film storage units holding 600 rolls each. Three printer/CRT workstations for the staff, and one public workstation are included. The workstations would have a computer terminal linked to the main computer to facilitate database searches. Each retrieval terminal would have a keyboard controller for user key entry of the desired frame location.

Manual key entry of frame location was considered more cost effective for this application. There are more expensive systems available which automatically link the search "hit" list to the film frame location, and automatically check a film code to verify that the correct roll is loaded prior to advancing the film. It was decided that since the user has to retrieve the desired film rolls manually, this costly automation would not add significantly to the efficiency of the retrieval process.

The microfilm reader-printers would advance the film from a manually key-entered frame location, electronically count (sense) the blip marks and stop at the correct frame. The workstations would have print-on-demand capability. The film roll would then be returned to the carousel film library.

The commercial marketplace offers more complex solutions to microform information retrievals, involving automated "picking" of the film roll and digital image scanning. The digital image can then be transported to any workstation in a network or linked through communications. This configuration would need only one film library, but is more costly and often needs to be custom designed for specific applications. Consequently, this complex technological approach was not further considered in this cost analysis.

D.4.2 Derivation of Costs

System costs include acquisition of a totally new micrographics system, with the equipment and services obtained through a systems integration contract. It assumes that the individual components are selected primarily on the basis of performance rather than the lowest price bid. The equipment items listed are sufficient to complete the conversion in three years. Estimated costs for a central computer system, software development, and systems integration are included.

The equipment prices listed are from GSA schedules when available. The duplicate microfilms are specified as 2,500-foot rolls for bulk loading of the printing equipment. The conversion personnel costs are for five operators and a supervisor, pegged at government salary scale rates of GS-4 and GS-6 respectively, with a 16% factor added for employee benefits. The original camera master films and the completed documents would be delivered to and kept at off-site storage.

There are one-time costs with a subtotal by line item, and also yearly cost items. Equipment maintenance is included at approximately 10% of original hardware costs for technical support contracts and repair parts. Costs for operating the existing reference system during the three year conversion effort, and long term off-site paper storage and preservation are included. Since this system utilizes computer-assisted index search and image retrievals, staff requirements subsequent to conversion are reduced. Three reference staff members at $19,200 per year could handle the daily reference workload.

Explanations of nomenclature and derivation of certain costs used in Table D-4 follow:

- High Speed Camera: A high speed microfilm camera with automated document transport system.
- Low Speed Camera: Fragile documents and refilms will be handled by a table top 16mm planetary camera.
- Index/QC CAR Units: Two workstations needed for camera master film inspection and index data entry.
- Processing Chemistry: Microfilm processing solutions based on daily use and replenishment; costs spread over three years of conversion.
- ANSI 16mm Film Mag: The positive microfilm prints inserted into ANSI film magazines for CAR use by staff and public; costs spread over three years needed to complete the conversion.
- Cam. Mstr. Film Rolls: Twelve million images require 2,280 rolls of 215-foot thin base microfilm, each roll containing 5,260 images; costs spread over three years of conversion.
- Dup. Film Pos. Rolls: 16mm duplicate microfilms for staff and public reference; costs spread over three years of conversion; replacement copies made in year seven.
- Dup. Film Neg. Rolls: 16mm duplicate negatives used to produce subsequent microfilm copies.
- Film Reel/Cartons: Boxes and plastic storage reels to store the duplicate negative microfilms; costs spread over three years.
- Cart.Ed./Loader: Cartridge editor/loader to load the ANSI film magazines.
- Staff & Pub. CAR Units: CAR microfilm retrieval workstations used by staff and walk-in public.

* Film Carrousel Units: Manual rotating microfilm storage units.

* Maintenance Exp./Yr.: Estimated (10% of original cost) yearly expenses for equipment and software maintenance and repair.

* Doc. Prep. Staff: The 8 million documents would require three GS-04 ($16,905/yr) for three years to prepare the records for conversion.

* Doc Prep Mat & Srvcs: $10,000 each for materials needed in document preparation and conservation laboratory services to repair damaged documents discovered during document preparation; costs spread over three years.

* Conversion Staff: Five staff members are needed: 2 camera operators, 1 film processor and duplicator, and 2 index/QC. A supervisor is also required. An additional quarter staff-year is required in year seven to make replacements for positive rolls used in staff and public reference.

* Reference Staff: Three full-time staff to handle user requests; workload phased in during first three years while conversion is partially completed.

* Master Film Storage: The camera master microfilms will be stored underground under archival conditions; mid-year costs used during first three years while conversion is partially completed.

* Dup. Neg. Storage: Direct duplicate negative microfilms require on-site storage; mid-year costs used during first three years while conversion is partially completed.

* Paper Storage: The converted documents would be transferred to low-cost underground storage; mid-year costs used during first three years while conversion is partially completed.

* Cur. Paper Syst. Costs: Three years of declining costs are included for paper records reference system operation; no further costs after conversion is completed.

* Cntrl. CPU, Index, Ntwrk & CRTs: Costs to obtain computer system, index magnetic storage, and related communication links and CRT equipment; replaced at year seven at 50% of year-one costs.

* System Integration: Contractor assistance to select, integrate, and install the system hardware and software.

* Sftwr (App./DBMS/Sys): Costs of licenses and customization services for application, index data base management system, and operating system software; replaced at year seven.

* Film Prsvtn/Inspct: Inspection program to ensure archival longevity of camera master microfilms; phased in during first three years.

* Training Materials: Development of a staff and public training program in use of the new system.

Ten-year cost figures for supporting this option are presented in Table D-4.

D.4.3 Advantages and Disadvantages

Advantages

- Computer-assisted retrieval systems provide information to requestors much faster than manual search systems.
- New production and retrieval equipment can provide higher quality images and hardcopy prints.
- Staff and public users are already familiar with microform systems.
- File integrity and storage space problems are addressed with a new film system.

Disadvantages

- There is an initial capital outlay for expensive production and retrieval equipment.
- Conversion to an automated search system is labor intensive.
- With this equipment configuration, electronic transmission of images to remote sites is not possible, requiring a decentralized system approach of multiple film copies at each site.
- Retrieval equipment is required at each remote search site.

CAR Microfilm System - Cost Breakdown (Actual Costs)

COMPONENT	QTY	EACH	YEAR 1	2	3	4	5	6	7	8	9	10	TOTAL
HIGH SPEED CAMERA	1	$104,500	$104,500										$104,500
LOW SPEED CAMERA	1	$4,740	$4,740										$4,740
INDEX/QC CAR UNITS	2	$10,500	$21,000										$21,000
PROCESSING CHEMISTRY	1	$18,450	$6,150	$6,150	$6,150				$1,538				$19,988
ANSI 16MM FILM MAG	4560	$0.59	$897	$897	$897								$2,690
CAM.MSTR.FILM ROLLS	2280	$8.41	$6,392	$6,392	$6,392								$19,175
DUP.FILM POS.ROLLS	400	$49.70	$6,627	$6,627	$6,627	$20	$20	$20	$19,880	$20	$20	$20	$39,879
DUP.FILM NEG.ROLLS	200	$60.22	$4,015	$4,015	$4,015								$12,044
FILM REELS/CARTONS	2280	$0.50	$380	$380	$380								$1,140
CART.ED./LOADER	1	$1,100	$1,100										$1,100
STAFF & PUB.CAR UNITS	4	$16,567	$66,268						$66,268				$132,536
FILM CARROUSEL UNIT	8	$450	$3,600										$3,600
MAINTENANCE EXP./YR			$21,024	$21,024	$21,024	$14,987	$18,300	$18,300	$12,487	$12,487	$12,487	$12,487	$164,606
DOCUMENT PREP STAFF	3	$16,905	$50,715	$50,715	$50,715								$152,145
DOC PREP SUP & SRVCS			$6,667	$6,667	$6,667								$20,000
CONVERSION STAFF													
[5-GS4 EMPL.]	5	$16,905	$84,525	$84,525	$84,525				$4,226				$257,801
[1-GS6 EMPL.]	1	$21,081	$21,081	$21,081	$21,081								$63,243
REFERENCE STAFF	3	$19,200	$9,600	$28,800	$48,000	$57,600	$57,600	$57,600	$57,600	$57,600	$57,600	$57,600	$489,600
MASTER FILM STORAGE	2280	$0.16	$61	$182	$304	$365	$365	$365	$365	$365	$365	$365	$3,101
DUP. NEG. STORAGE	2280	$0.18	$68	$205	$342	$410	$410	$410	$410	$410	$410	$410	$3,488
PAPER STORAGE			$873	$2,620	$4,367	$5,240	$5,240	$5,240	$5,240	$5,240	$5,240	$5,240	$44,540
CUR. PAPER SYST. COSTS			$86,667	$52,000	$17,333								$156,000
CNTRL.CPU,INDEX STOR. NTWRK, & CRTS	1	$50,000	$50,000						$25,000				$75,000
SYSTEM INTEGRATION	1	$40,000	$40,000						$40,000				$80,000
SFTWR [APP./DBMS/SYS]	1	$30,000	$30,000						$30,000				$60,000
FILM PRSVTN/INSPCT.	2280	$0.16	$61	$182	$304	$365	$365	$365	$365	$365	$365	$365	$3,102
TRAINING MATERIALS			$5,000										$5,000
TOTALS			$632,009	$292,461	$279,121	$78,987	$82,300	$82,300	$263,379	$76,487	$76,487	$76,487	$1,940,018

Table D-4

D.5 Digital Image/Optical Disk System

This system includes conversion of documents to electronic images similar to the ODISS project. Production steps of indexing, quality review, optical disk storage, and workstation reference are included. All production steps to convert the twelve million images would be performed by in-house staff.

D.5.1 Description

The record holdings would be prepared for scanning by in-house staff. The conversion would begin with a high speed scanner to capture the 75 percent of the documents that are high-speed-transportable. The remaining documents would be converted using a tabletop scanner. Three workstations would be used to perform index data entry and image quality control. The images would be stored on twelve-inch, write once (WORM) optical disks located in three automated jukeboxes. The images would be retrievable using high resolution monitors, and high quality laser printers would be used to replicate images back to paper.

Paper Scanners

A high speed scanner would capture two-sided documents on one pass to save labor, processing time, and wear and tear on the documents. It must not damage the original paper in any way. In order to convert the documents within the allotted time frame of 5250 hours (three years), the scanner must be able to process 1,143 pages per hour or 19 pages per minute. With 50% of the pages double-sided, it would require the capture of 1,715 images per hour. This rate is within the capabilities of currently available high speed scanners. A tabletop scanner could convert the two million documents unsuitable for high speed scanning within three years, which equates to 571 images per hour or 6.3 seconds per image.

Index and Quality Control Workstations

Index data would be key-entered for each file. Assuming sixty seconds for each of the 400,000 files, it would take one operator 6,667 hours to complete the task. In comparison, two workstations with two operators only need to index one file every two minutes. This system assumes a conservative approach of key entry of index data. Depending on the actual application, automated input of these data by the use of optical character recognition (OCR) or other means might be possible in the future.

Twenty percent of the captured files are inspected for index and image quality, requiring 1,333 hours to complete at 60 seconds per file. At this rate, 80,000 files would require 5,250 hours, or one file every four minutes.

Optical and Magnetic Storage

Twelve-inch write-once optical disks would be used to store the digital images. Each disk would contain 4.4 gigabytes of user data, and store 80,000 images. Three jukeboxes containing 50 disks each would be required to hold the estimated 150 disks. Since each jukebox has retrieval robotics, system response would be consistently fast. Digital images would be copied to non-erasable optical disks, while index data would be stored on magnetic disks. Three hundred disks would be acquired in order to produce both a primary and a backup copy.

D.5.2 Derivation of Costs

System costs are for the integration of a totally new system based upon digital image and optical disk technologies, with the government obtaining the equipment and services through a systems integration contract. The in-house conversion would require three years, with costs for maintaining the current paper system declining at a rate of 33 percent per year. Estimated costs for a central computer system, workstation integration, and software development are provided. Equipment maintenance was estimated to be 10% of original equipment cost per year.

The component prices used were taken from prices for off-the-shelf units that do not require extensive customization. Conversion personnel costs are for five operators and one supervisor, pegged at federal GS-4 and GS-6 pay scales respectively, with a 16% factor added for employee benefits. After the conversion is completed, reference services would be handled by a single staff member and the conversion staff would be released to other duties. Only the reference equipment and software need to be replaced at the end of the seven year life cycle.

Explanations of nomenclature and derivation of certain costs used in Table D-5 follow:

- ☼ <u>High Speed Scanner w/Enhancement</u>: Scanner with document transport system and automatic image enhancement.

- ☼ <u>Low Speed Scanner w/Enhancement</u>: Tabletop, hand-fed scanner for fragile documents and rescans.

- ☼ <u>Input Workstations</u>: Digital image workstations with high resolution displays, indexing and quality control software.

- ☼ <u>Control CPU, Image Buffer, Index Storage & Network</u>: Costs to obtain computer system, image buffers, index magnetic storage, related communication links, and CRT equipment.

- ☼ <u>Optical Disk Jukebox</u>: Optical disk library device for automatically storing and retrieving up to 50 optical disks, with two optical drives; three jukeboxes are required; replaced with then-current technology at year seven.

- ☼ <u>Optical Disk Media</u>: 300 twelve-inch write-once, read-many-times (WORM) disks; replaced at year seven.

- ☼ <u>Retrieval Wkstns</u>: Digital image workstations with high resolution displays and retrieval software; replaced at year seven.

- ☼ <u>System Integration</u>: Contractor assistance to select, integrate, and install the system hardware, software, and communications; also needed in year seven to assist with system replacement.

- ☼ <u>Retrieval Staff</u>: One full-time staff (i.e., workload of one staff-year) to handle user requests; phased in during conversion period of first three years.

- ☼ <u>Image Printers</u>: Digital image laser printers with servers to produce high quality image and index data hardcopies; replaced in year seven.

- Software: Application, index data base management system, and operating system software licenses and customization services.

- Document Preparation Personnel: The eight million documents would require three staff, pegged at a federal GS-04 pay scale ($16,905/yr), for three years to prepare the records for conversion.

- Document Prep Supplies & Srvcs: $10,000 each for materials needed in document preparation and conservation laboratory services to repair damaged documents discovered during document preparation; costs spread over three years.

- Conversion Personnel: Five staff members are needed: two scanner operators, two index/quality control technicians, and one to perform database management functions. A supervisor to serve as system manager is also required.

- Current Paper Syst: Three years of declining costs for operation of paper records reference system; phased out as conversion moves toward completion.

- Training Materials: Staff and public training program development costs.

- System Maintenance: Estimated yearly expenses (10% of original costs) for equipment and software maintenance and repair.

- Paper Storage: Converted documents would be transferred to low-cost underground storage; phased in during three-year conversion period.

Ten-year cost figures for supporting this option are presented in Table D-5.

D.5.3 Advantages and Disadvantages

Advantages

- Digital imaging offers image processing techniques to enhance the visual quality of the captured images.
- Computer indexing offers rapid search of the database to find documents of interest.
- Jukebox storage of optical disks permits rapid retrieval of electronic images from among massive quantities.
- Once captured digitally, identical copies of the image files can be created with no loss of data. There should never be any need to rescan the original documents.

Disadvantages

- Digital images and related indexes are in electronic form and cannot be accessed or made usable without the use of computer equipment and software.
- The complex hardware used in optical digital imagery systems require ongoing service contracts for the life of the system.
- Optical disk system conversions generally require large commitments of funds in the early years to acquire image capture and retrieval hardware and software.

Digital Image/Optical Disk System - Cost Breakdown (Actual Costs)

COMPONENT	QTY	EACH	YEAR 1	2	3	4	5	6	7	8	9	10	TOTAL
HIGH SPEED SCANNER W/ENHANCEMENT	1	$75,000	$75,000										$75,000
LOW SPEED SCANNER W/ENHANCEMENT	1	$10,000	$10,000										$10,000
INPUT WORKSTATIONS	3	$15,000	$45,000										$45,000
(2 Index, 1 OC)													
CONTROL CPU, IMAGE BUFFER,			$100,000						$50,000				$150,000
INDEX STORAGE & NETWORK													
OPTICAL DISK JUKEBOX	3	$75,000	$225,000						$112,500				$337,500
(w/2 drives each &													
w/capacity of 50 disks each)													
OPTICAL DISK MEDIA	300	$300	$30,000	$30,000	$30,000				$45,000				$135,000
(4.4 GB ea. 80K images ea.)													
RETRIEVAL WKSTNS	2	$15,000	$30,000						$15,000				$45,000
SYSTEM INTEGRATION			$75,000						$75,000				$150,000
RETRIEVAL STAFF	1	$19,200	$3,200	$9,600	$16,000	$19,200	$19,200	$19,200	$19,200	$19,200	$19,200	$19,200	$163,200
IMAGE PRINTERS	2	$10,000	$20,000						$10,000				$30,000
SOFTWARE			$50,000						$60,000				$110,000
(APPLICATION/DBMS/SYSTEM)													
DOCUMENT PREPARATION PERSONNEL	3	$16,905	$50,715	$50,715	$50,715								$152,145
DOCUMENT PREP SUPPLIES & SRVCS			$6,667	$6,667	$6,667								$20,000
CONVERSION PERSONNEL													
[5-GS4 EMPL.]	5	$16,905	$84,525	$84,525	$84,525								$253,575
[1-GS6 EMPL.]	1	$21,081	$21,081	$21,081	$21,081								$63,243
CURRENT PAPER SYST. COSTS			$86,667	$52,000	$17,333								$156,000
TRAINING MATERIALS			$5,000										$5,000
SYSTEM MAINTENANCE /YR			$55,500	$55,500	$55,500	$42,500	$47,000	$47,000	$24,750	$24,750	$24,750	$24,750	$402,000
PAPER STORAGE			$873	$2,620	$4,367	$5,240	$5,240	$5,240	$5,240	$5,240	$5,240	$5,240	$44,540
TOTALS			$974,228	$312,708	$286,188	$66,940	$71,440	$71,440	$416,690	$49,190	$49,190	$49,190	$2,347,203

Table D-5

D.6 Upgraded (CAR) Microfilm System Using Service Bureau Conversion

This alternative is identical to the one presented in section D.4 except that a commercial service bureau would perform all tasks involved in the conversion of the paper documents to microfilm suitable for use in a CAR system.

D.6.1 Description

The service bureau would be under contract to complete the conversion within three years. Service bureau duties would include document preparation, microfilming, indexing, quality control, and film duplication. A high quality automated transport, non-rotary camera would create high resolution images. The output media would be 16mm silver halide, 215-foot thin film blip-encoded camera masters. Two positive user copies, and one negative printing master would be produced. The service bureau would provide all required production equipment, staff, and supplies. A staff member would provide conversion oversight and monitor the contractor's progress and quality performance levels. The service bureau would also be contracted to produce replacements for the staff and public reference copies of the microfilms in year seven.

D.6.2 Derivation of Costs

The major cost of service bureau selection would be the expenditure during the first three years for services to convert the records from paper to microfilm. Cost estimates were obtained from quotes provided by established commercial service bureaus. Since no government-owned conversion hardware would be required, the only costs would be for CAR retrieval terminals, equipment maintenance, paper conservation, off-site master film and paper storage, and retrieval computer software and hardware integration services. An ongoing annual cost of $19,200 would be for a single retrieval staff member to support reference activities.

Explanations of nomenclature and derivation of certain costs used in Table D-6 follow:

- Conversion Serv. Bur: Three years of service contractor costs for document preparation, microfilming, indexing, quality control, and duplicating.

- Conversion Oversight: One in-house staff member for three years to monitor the service bureau's performance.

- Staff/Pub. CAR Units: CAR microfilm retrieval workstations used by staff and walk-in public.

- Film Carrousel Units: Manual rotating microfilm storage units.

- Equipment Maintenance: Yearly expenses (10% of original costs) for equipment and software maintenance and repair.

- Reference Staff: Three full-time staff to handle user requests; phased in during three years needed to complete the conversion.

- Conservation Srvcs: $10,000 for conservation laboratory services to repair damaged documents discovered during document preparation; costs spread over three years.

- Master Film Storage: Camera master microfilms will be stored underground under archival conditions; phased in during three years needed to complete the conversion.

- Dup. Neg. Stor: Duplicate negative microfilms require on-site storage; phased in during three years needed to complete the conversion.

- Paper Storage: Converted documents transferred to low-cost underground storage; phased in during three years needed to complete the conversion.

- Dup. Film-Pos. Rolls: Staff and public positive microfilm replacements produced at year seven by a service bureau.

- Current Paper Costs: Three years of declining costs for paper records reference system operation as it is phased out.

- Cntrl CPU/Index Stor. Network & CRT's: Costs to obtain computer system, index magnetic storage, related communication links, and CRT equipment; replaced at year seven at 50% of year-one costs.

- System Integration: Contractor assistance to select, integrate, and install system hardware and software.

- Software (App./DBMS/Sys): Estimated costs of licenses and customization services for application, index data base management system, and operating system software; replaced at year seven.

- Film Prsrvtn/Inspct: Camera master rolls require a routine inspection program to ensure archival longevity; phased in during first three years.

- Training Materials: Staff and public training program development costs.

Ten-year cost figures for supporting this option are presented in Table D-6.

D.6.3 Advantages and Disadvantages

Advantages

- Computer-assisted retrieval systems provide information to requestors much faster than annual search systems.
- New production and retrieval equipment can provide higher quality images and hardcopy prints.
- Staff and public users are already familiar with microform systems.
- File integrity and storage space problems are addressed with a new film system.
- There are no capital costs for new hardware or staff to accomplish the conversion.
- The service bureau would be under contractual obligations to complete the conversion according to a pre-established time schedule.
- Service bureaus generally use the latest equipment and software technologies available; government-owned systems begin to become obsolete from the time of acquisition.

Disadvantages

- There is an initial capital outlay for conversion services and retrieval equipment.
- Transmission of images to remote sites is not possible, requiring a decentralized system approach of multiple film copies at each site.
- Retrieval equipment is required at each remote search site.
- Archival documents would be handled by non-archives service bureau staff.
- Service bureau contracts may be more costly due to profit margins and short turnaround times required by the archival institution for the conversion.

CAR Microfilm System with Conversion by Service Bureau
Cost Breakdown
(Actual Costs)

COMPONENT	QTY	EACH	YEAR 1	2	3	4	5	6	7	8	9	10	TOTAL
CONVERSION SERV.BUR.			$666,667	$666,667	$666,667								$2,000,001
CONVERSION OVERSIGHT	1	$19,200	$19,200	$19,200	$19,200								$57,600
STAFF/PUB.CAR UNITS	4	$16,567	$66,268						$66,268				$132,536
FILM CARROUSEL UNITS	8	$450	$3,600										$3,600
EQUIPMENT MAINTENANCE			$14,987	$14,987	$14,987	$14,987	$14,987	$14,987	$12,487	$12,487	$12,487	$12,487	$139,868
REFERENCE STAFF	3	$19,200	$9,600	$28,800	$48,000	$57,600	$57,600	$57,600	$57,600	$57,600	$57,600	$57,600	$489,600
CONSERVATION SRVCS			$3,333	$3,333	$3,333								$10,000
MASTER FILM STORAGE	2280	$0.16	$61	$182	$304	$365	$365	$365	$365	$365	$365	$365	$3,101
DUP.NEG.STOR.	2280	$0.18	$68	$205	$342	$410	$410	$410	$410	$410	$410	$410	$3,486
PAPER STORAGE			$873	$2,620	$4,367	$5,240	$5,240	$5,240	$5,240	$5,240	$5,240	$5,240	$44,540
DUP.FILM-POS.ROLLS	4560	$18.62							$84,907				$84,907
CURRENT PAPER COSTS			$86,667	$52,000	$17,333								$156,000
CNTRL CPU/INDEX STOR NETWORK, & CRT'S	1	$50,000	$50,000						$25,000				$75,000
SYSTEM INTEGRATION	1	$40,000	$40,000						$40,000				$80,000
SOFTWARE [APPLIC./DBMS/SYS]			$30,000						$30,000				$60,000
FILM PRSRVTN/INSPCT.	2280	$0.16	$61	$183	$304	$365	$365	$365	$365	$365	$365	$365	$3,102
TRAINING MATERIALS			$5,000										$5,000
TOTALS			$996,385	$788,177	$774,837	$78,967	$78,967	$78,967	$322,642	$76,467	$76,467	$76,467	$3,348,341

Table D-6

D.7 Digital Image/Optical Disk System Using Service Bureau Conversion

This alternative is identical to the one in section D.5 except that a commercial service bureau would be employed to perform all tasks involved in the conversion of the paper documents to digital images and storing the images on optical disk.

D.7.1 Description

A commercial service bureau would be contracted to complete the conversion within three years. The contractor would be responsible for all production including minor document preparation, scanning, indexing, quality control, and storage of document images on optical disks. An in-house staff member would be designated as a conversion oversight monitor to observe and measure the contractor's performance.

In the case of digital image capture, a resolution of 200 lines per inch would be used. Output for the retrieval system configuration would be identical to that described in Section D.5.1. Twelve-inch disks with a capacity of 4.4 gigabytes per disk would be used to write one primary and one security backup copy.

D.7.2 Derivation of Costs

As for the microfilm conversion using a service bureau, the primary expense will be for the conversion itself. These costs eliminate the need to acquire staff and hardware for converting the document holdings. Additional expenditures are for retrieval computer hardware and software integration, system maintenance, and paper conservation. A jukebox network of three devices, each with two optical drives would be required to service reference requests. (Optical disk media would be provided by the contractor and are included in the service bureau cost estimates.) Off-site paper storage, paper preservation expense, and system supplies are other inherent expenditures. Equipment maintenance costs are estimated at 10% per year based upon original hardware and software costs for the reference system.

Explanations of nomenclature and derivation of certain costs used in Table D-7 follow:

- Conversion Serv. Bur: Service bureau contractor costs for document preparation, scanning, indexing, quality control, and optical disk supplies.

- Cntrl CPU, Img Buf, Ind Stor & Ntwrk: Costs to obtain computer system, image servers, index magnetic storage, related communication links, and CRT equipment for the retrieval system.

- Optical Disk Jukebox: Optical disk library for storing and retrieving up to 50 optical disks, with two optical drives; three jukeboxes are required at $75,000 each; replaced at year seven at 50% of Year 1 costs.

- OD Media + Labor: Replacement costs at seven years for the optical media; includes service bureau labor to perform digital copying and/or conversion.

- Retrvl Wkstns: Three digital image workstations with high resolution displays and retrieval software at $15,000; replaced at year seven at 50% of Year 1 cost.

- System Integration: Contractor assistance to select, integrate, and install the reference system hardware, software, and communications.

- Retrieval Staff: One full-time staff (i.e., workload of one staff-year) to handle user requests; phased in over the three-year conversion period.

- Image Printers: Digital image laser printers with servers to produce high quality prints of image and index data; replaced at year seven.

- Software: Cost of licenses and customization services for application, index data base management system, and operating system software; costs incurred again in year seven for replacement of the retrieval system.

- Conversion Oversight: One in-house staff member for three years to monitor the service bureau's performance.

- Syst Maint./Yr: Estimated yearly expenses (10% of original costs) for equipment and software maintenance and repair.

- Cur Syst. Paper Costs: Three years of declining costs for paper records reference system operation; phased out as conversion is completed.

- Conservation Srvcs: $10,000 for conservation laboratory services to repair damaged documents discovered during document preparation; costs spread over three years.

- Training Materials: Staff and public training program development costs.

- Paper Storage: Converted documents would be transferred to low-cost underground storage; phased in over three years.

Ten-year cost figures for supporting this option are presented in Table D-7.

D.7.3 Advantages and Disadvantages

Advantages

- Digital imaging offers image processing techniques to enhance the visual quality of the captured images.
- Computer indexing offers rapid search of the database to find documents of interest.
- Jukebox storage of optical disks permits rapid retrieval of electronic images from among massive quantities.
- Once captured digitally, identical copies can be created with no loss of data.
- Service bureau utilization eliminates the need to obtain expensive conversion equipment.
- Service bureau would be under contract to deliver the completed products within the specified time.
- Need for skilled technicians to operate the conversion equipment is eliminated.

Disadvantages

- Digital images and related indexes are in electronic form and cannot be accessed or made usable without the use of computer equipment and software.
- The complex hardware used in optical digital imagery systems require ongoing service contracts for the life of the system.
- Service bureau contracts may be more costly due to profit margins and short turnaround times required by the archival institution for the conversion.
- Archival documents would be handled by non-archives service bureau staff.
- Increased product quality monitoring would be needed to validate contractor's performance level.

Digital Image/Optical Disk System with Conversion by Service Bureau - Cost Breakdown (Actual Costs)

COMPONENT	QTY	EACH	YEAR 1	2	3	4	5	6	7	8	9	10	TOTAL
CONVERSION SERV.BUR. (OVER 3 YR.PERIOD)			$666,667	$666,667	$666,667								$2,000,001
CNTRL.CPU,IMG.BUF. IND.STOR & NTWRK			$100,000						$50,000				$150,000
OPTICAL DISK JUKEBOX [W/2 drives each & W/50 DISKS]	3	$75,000	$225,000						$112,500				$337,500
OD MEDIA+LABOR									$58,000				$58,000
RETRVL. WKSTNS	2	$15,000	$30,000						$15,000				$45,000
SYSTEM INTEGRATION			$67,000						$67,000				$134,000
RETRIEVAL STAFF	1	$19,200	$3,200	$9,600	$16,000	$19,200	$19,200	$19,200	$19,200	$19,200	$19,200	$19,200	$163,200
IMAGE PRINTERS	1	$10,000	$10,000						$5,000				$15,000
SOFTWARE [APPL.,DB,SYST.]			$50,000						$50,000				$100,000
CONVERSION OVERSIGHT [1 GS06]	1	$19,200	$19,200	$19,200	$19,200								$57,600
SYST.MAINT./YR			$41,500	$41,500	$41,500	$41,500	$41,500	$41,500	$23,250	$23,250	$23,250	$23,250	$342,000
CUR.SYST.PAPER COSTS			$86,667	$52,000	$17,333								$156,000
CONSERVATION SRVCS			$3,333	$3,333	$3,333								$10,000
TRAINING MATERIALS			$5,000										$5,000
PAPER STORAGE			$873	$2,620	$4,367	$5,240	$5,240	$5,240	$5,240	$5,240	$5,240	$5,240	$44,540
TOTALS			$1,308,440	$794,920	$768,400	$65,940	$65,940	$65,940	$405,190	$47,690	$47,690	$47,690	$3,617,841

Table D-7

D.8 Comparison of Costs of System Alternatives

In order to show how the costs of the various alternatives compare to each other, Table D-8 presents a cumulative cost comparison chart which shows the related costs each of the alternatives over a ten-year period. Cumulative cost totals are also shown for each year, utilizing the *net present value* method based upon a ten percent discount rate as prescribed by OMB Circular A-94. Present value is a concept used to equate future dollars with present dollars and to compare costs that occur at different points in the future. Simply put, a dollar given to you a year in the future is not worth the same as a dollar given to you today; a dollar given to you two years in the future is worth still less, and so on.

For the six alternatives, Table D-8 shows the actual yearly costs brought forward from Table D-2 through Table D-7. Then for each alternative, the appropriate present value discount factor is applied to the actual cost figure for each year to arrive at a discounted yearly cost. Two other rows are then computed, one showing cumulative actual costs over a ten-year period, and the other showing cumulative discounted costs (from the application of present value discount factor) over the same period.

In further explanation of the application of the present value discount factor, the recurring annual costs of the paper-based alternative amount to $128,513 for the first three years *as the money is valued at the outset, i.e, on day one of year one.* At the end of the third year, *in terms of constant dollars*, the cumulative costs will have been three times $128,513 or $385,539. Nevertheless, because of inflation and other economic factors, the real value of that $385,539 would be only $351,483 *in terms of the same year-one, day-one dollars.* Use of present value points out the disadvantage of solutions which require significant capital outlays for equipment or services in the early years as compared to solutions where costs may be deferred or spread out over a number of years during which the value of money declines.

Figure D-1 is a graphical representation of Table D-8 showing the cumulative discounted costs over a ten-year period.

Cumulative Cost Comparison

	YEAR 1	2	3	4	5	6	7	8	9	10
PAPER:										
YEARLY COST	$128,513	$128,513	$128,513	$104,000	$104,000	$104,000	$104,000	$104,000	$104,000	$104,000
PV DISC.FACTOR	1.000	0.909	0.826	0.751	0.683	0.621	0.564	0.513	0.467	0.424
DISC.YRLY COST	$128,513	$116,818	$106,152	$78,104	$71,032	$64,584	$58,656	$53,352	$48,568	$44,096
CUM. ACTUAL	$128,513	$257,026	$385,539	$489,539	$593,539	$697,539	$801,539	$905,539	$1,009,539	$1,113,539
CUM. DISC.	$128,513	$245,331	$351,483	$429,587	$500,619	$565,203	$623,859	$677,211	$725,779	$769,875
MANUAL FILM:										
YEARLY COST	$364,700	$299,828	$286,908	$70,455	$70,455	$70,455	$122,407	$70,455	$70,455	$70,455
PV DISC.FACTOR	1.000	0.909	0.826	0.751	0.683	0.621	0.564	0.513	0.467	0.424
DISC.YRLY COST	$364,700	$272,543	$236,986	$52,912	$48,121	$43,753	$69,038	$36,143	$32,903	$29,873
CUM. ACTUAL	$364,700	$664,527	$951,435	$1,021,890	$1,092,345	$1,162,800	$1,285,207	$1,355,662	$1,426,117	$1,496,572
CUM. DISC.	$364,700	$637,243	$874,228	$927,140	$975,261	$1,019,014	$1,088,051	$1,124,195	$1,157,097	$1,186,970
CAR MICROFILM:										
YEARLY COST	$632,009	$292,461	$279,121	$78,987	$82,300	$82,300	$263,379	$76,487	$76,487	$76,487
PV DISC.FACTOR	1.000	0.909	0.826	0.751	0.683	0.621	0.564	0.513	0.467	0.424
DISC.YRLY COST	$632,009	$265,847	$230,554	$59,319	$56,211	$51,108	$148,545	$39,238	$35,719	$32,430
CUM. ACTUAL	$632,009	$924,471	$1,203,592	$1,282,579	$1,364,879	$1,447,179	$1,710,558	$1,787,045	$1,863,532	$1,940,018
CUM. DISC.	$632,009	$897,857	$1,128,411	$1,187,730	$1,243,941	$1,295,050	$1,443,595	$1,482,833	$1,518,552	$1,550,983
DI/CO:										
YEARLY COST	$974,228	$312,708	$286,188	$66,940	$71,440	$71,440	$416,690	$49,190	$49,190	$49,190
PV DISC.FACTOR	1.000	0.909	0.826	0.751	0.683	0.621	0.564	0.513	0.467	0.424
DISC.YRLY COST	$974,228	$284,251	$236,391	$50,272	$48,794	$44,364	$235,013	$25,234	$22,972	$20,857
CUM. ACTUAL	$974,228	$1,286,935	$1,573,123	$1,640,063	$1,711,503	$1,782,943	$2,199,633	$2,248,823	$2,298,013	$2,347,203
CUM. DISC.	$974,228	$1,258,479	$1,494,870	$1,545,142	$1,593,935	$1,638,300	$1,873,313	$1,898,547	$1,921,519	$1,942,376
CAR S/B:										
YEARLY COST	$996,385	$788,177	$774,837	$78,967	$78,967	$78,967	$322,642	$76,467	$76,467	$76,467
PV DISC.FACTOR	1.000	0.909	0.826	0.751	0.683	0.621	0.564	0.513	0.467	0.424
DISC.YRLY COST	$996,385	$716,453	$640,015	$59,304	$53,934	$49,038	$181,970	$39,227	$35,710	$32,422
CUM. ACTUAL	$996,385	$1,784,562	$2,559,400	$2,638,367	$2,717,333	$2,796,300	$3,118,942	$3,195,408	$3,271,875	$3,348,341
CUM. DISC.	$996,385	$1,712,838	$2,352,854	$2,412,158	$2,466,092	$2,515,130	$2,697,100	$2,736,328	$2,772,038	$2,804,459
DI/CO SB:										
YEARLY COST	$1,308,440	$794,920	$768,400	$65,940	$65,940	$65,940	$405,190	$47,690	$47,690	$47,690
PV DISC.FACTOR	1.000	0.909	0.826	0.751	0.683	0.621	0.564	0.513	0.467	0.424
DISC.YRLY COST	$1,308,440	$722,583	$634,699	$49,521	$45,037	$40,949	$228,527	$24,465	$22,271	$20,221
CUM. ACTUAL	$1,308,440	$2,103,361	$2,871,761	$2,937,701	$3,003,641	$3,069,581	$3,474,771	$3,522,461	$3,570,151	$3,617,841
CUM. DISC.	$1,308,440	$2,031,023	$2,665,722	$2,715,243	$2,760,280	$2,801,228	$3,029,755	$3,054,220	$3,076,492	$3,096,712

Table D-8

Cumulative Cost Graph

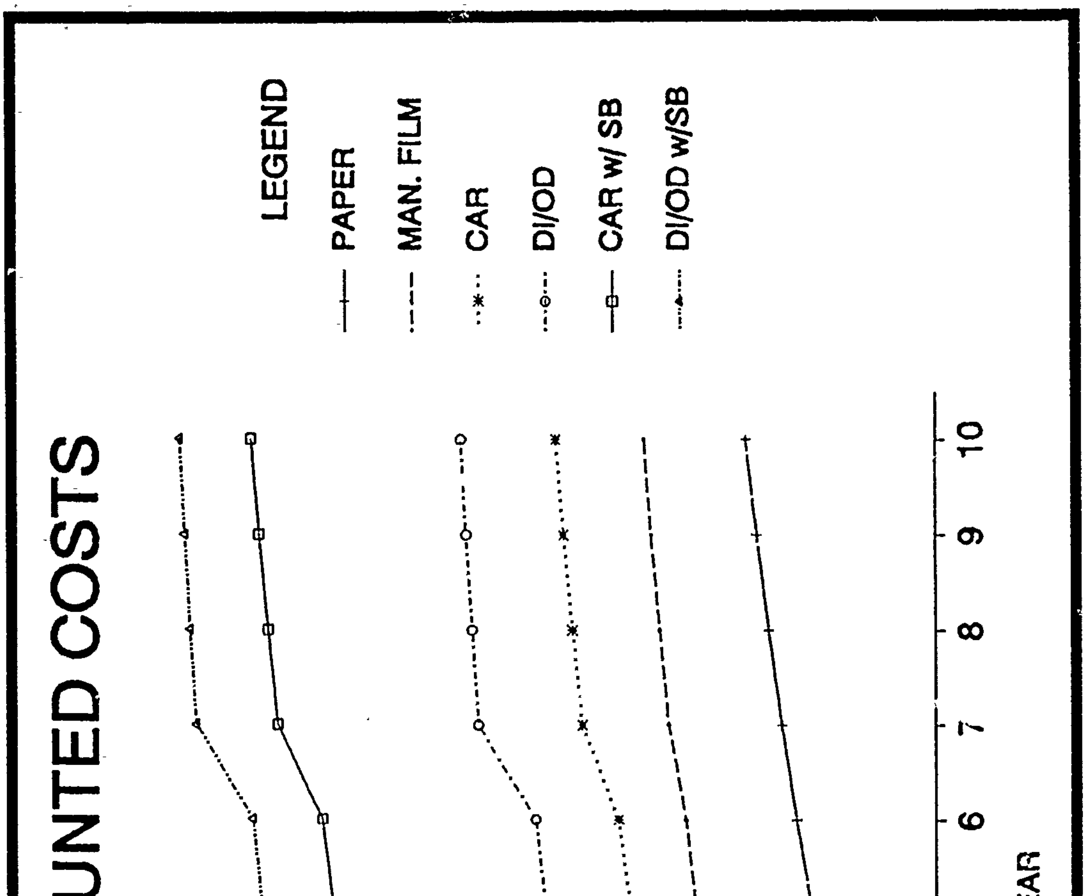


Figure D-1

D.9 Interpretation of the Model

As noted earlier, the purpose behind this appendix is to present a cost comparison of various records conversion alternatives with a baseline consisting of maintaining a set of original paper records and providing reference services directly from those original records. In order to standardize the analysis, a model consisting of a generic, hypothetical set of holdings was used.

In viewing Figure D-1, the plot for the baseline (paper) option appears at the bottom of the graph. None of the plots for the five alternatives involving a conversion intersects the plot for the baseline option. From this, one can infer that for the model collection under the parameters selected, conversion of the paper records to any of the alternate forms evaluated cannot be justified on the basis of cost alone over the first ten years. Nor are any sharp trends discernable in the plots that would suggest a cost payback would be realized within the near future beyond ten years.

In stating this conclusion, it is important to reemphasize that the plots in Figure D-1 reflect the costs associated with those parameters selected for the model and are not universally applicable to other document holdings with different numbers or attributes. Furthermore, differences in operational factors would influence the model. For example, if reference requirements included availability at more than one physical location, separate or networked reference systems would be required at the individual sites, thereby increasing the costs of some options. If labor and records storage costs differed radically from those used in the model, then substitutions would have to be made to reflect local conditions.

The model could also be influenced by technological evolution and market forces. For example, from current trends, it is anticipated that the development of optical and other storage technologies will progress to the point where their costs will be relatively inconsequential in terms of current dollars. In other words, over the next decade, advances in the technologies used to store massive amounts of digital data will probably not require expensive mechanical drives and will probably have increased capacities perhaps a thousand-fold greater than the best current storage media. Although the cost tables for the two optical disk options (Table D-5 and Table D-7) show system replacements in the seventh year which involve replacement of the optical disk systems and optical disks with new generations of the same technology, it is quite conceivable that an entirely new and much less expensive technology might be available that would radically reduce costs and thus commensurately affect the cost model plots.

In the final analysis, the comparisons provided by the model are intended to act only as a guide to show the many factors that influence the relative costs of conversion to and usage of alternative systems. The charts and graph cannot and do not cover all situations.

APPENDIX E

DATA COLLECTION FORMS

APPENDIX E. DATA COLLECTION FORMS

Indexing Input Data Collection Form

Date of interview or of filling out this form ______________

Person interviewed or filling out this form ______________

Interviewer (if done as an interview) ______________

1. Please tell us how easy or hard it was for you to learn how to use the ODISS indexing station:

   a. How easy or hard was it to learn to operate the index work station on a scale of 1 to 10 (1 = easiest & 10 = hardest)?

      ______

   b. What term most closely describes how easy or hard it was to learn to operate the index station? Check only one.

      Very easy ___

      Somewhat easy ____

      Average ______

      Somewhat difficult _______

      Very difficult ________

   c. What were the hardest things to learn in operating the station?

2. After you have learned to operate the station, how well overall does the ODISS index station work on a scale of 1 to 10 (1 = lowest rating & 10 = highest rating)? ________

Figure E-1, Page 1 of 3

3. Is the writing/printing on the image of the CMSR file jacket easy to read on the screen?

   a. Always easy ____

   b. Usually easy, occasionally hard ____

   c. Often difficult ____

4. When the jacket is hard to read, is this most often due to

   a. Poor image quality ______

   b. Illegible writing on the original document _____

   c. Not sure of the cause _______

5. Are the code tables easy to use? a. Yes ___ b. No ___

   If your answer is "No," which tables cause problems?

   c. All ___ d. Regiment ___ e. Company ___ f. Rank ___

   What are the problems? Check all that apply:

   g. Retrieval of a table too slow ___

   h. Scrolling within a table too slow ___

   i. Code numbers hard to remember ___

   j. Other ___ Describe the problem briefly:

6. Is it easy to use default settings when indexing files?

   a. Yes ___ b.No ___

   If "No," please describe the problem briefly:

Figure E-1, Page 2 of 3

7. Are the function keys easy to use? a. Yes ___ b. No ___

If "No," which keys cause the problems? Check c or d:

c. All ___

d. A specific key or keys ___ List the problem key(s):

What are the problems? Check all that apply:

e. Keys not well identified ___

f. Keys not well placed ___

g. Too many to learn & remember easily ___

h. Other ___ Describe problem briefly:

8. Any other problems with indexing input?

9. Any especially good features of the current indexing input?

10. Any suggestions for improving indexing input?

11. If you could change only one thing at the index station, what would it be?

Figure E-1, Page 3 of 3

Standard Procedures For Indexing

1. When there is no middle initial use NMI, no first initial use NFI, and no last name NLI.
2. When the rank or company name has been omitted, it should be noted under remarks.
3. In the remarks fields, key in the exact headings as noted, SEE ALSO: or CARDS FILED WITH: using colons following each.
4. Include under remarks any information written on the jacket, including information noted in the name fields.
5. When there is no information to be entered under remarks indicate by using NONE.
6. Do not use any other punctuation, except when it is a part of the name. Examples: O'Donnell, Cramer-Thomas

   Use a comma between the surname and the first initial or first name in the remarks fields. Examples: Smith, John R or West, B W
7. Jr or Sr are keyed in the middle initial field in the regular indexing fields. If there is a middle initial, space, then include without periods in the regular indexing fields and under remarks.
8. The use of first (1st) and second (2nd) in the regular fields will require that you use roman numerals instead. However, first (1st) and second (2nd) are acceptable in the remarks fields.
9. When a surname has two names, leave a space when a space is indicated on the jacket.
10. Names beginning with Mc are entered as one word. Example: McDonald
11. Extra company names are entered under remarks.
12. Zeros are always entered where no rank or company is given for these fields.
13. When keying in the information under remarks, enter it exactly as it is noted.

Figure E-2, Page 1 of 2

14. Multiple middle initials or middle names are all entered in the middle initial fields followed by a space.

15. When a jacket does not appear or is not legible at the index station, the Index Operator should use the regimental cards for abstracting the necessary information. It is the responsibility of the Quality Control Operator to ensure that the information corresponds to the information on the jacket. It is not required that the Indexer pull the jacket.

16. Any company names that are not coded should be bought to the attention of the supervisor.

17. Information in the name and the remarks fields must be entered in the first prompt field.
Example: /Fields not /-Fields

Figure E-2, Page 2 of 2

Quality Control Data Collection Form

Date of interview or of filling out this form ______________

Person interviewed or filling out this form ______________

Interviewer (if done as an interview) ______________________

1. Please tell how easy or hard it was for you to learn how to use the ODISS quality control station:

   a. How easy or hard was it to learn to operate the QC work station on a scale of 1 to 10 (1 = easiest & 10 = hardest)?

   ______

   b. What term most closely describes how easy or hard it was to learn to operate the QC station? Check only one.

   Very easy ___

   Somewhat easy ____

   Average ______

   Somewhat difficult _______

   Very difficult _________

   c. What were the hardest things to learn in operating the station?

2. After you have learned to operate the station, how well overall does the ODISS quality control station work on a scale of 1 to 10 (1 = lowest rating & 10 = highest rating)?

   ____

3. Is the index record for each file easy to read on the screen?
   a. Yes ____ b. No ____

4. Is it easy to correct indexing mistakes?

   a. Yes ____ b. No ____ If "No," describe the problem:

Figure E-3, Page 1 of 2

5. Are the function keys easy to use? a. Yes ___ b. No ___

If "No," which keys cause the problems? Check c or d:

c. All ___

d. A specific key or keys ___ List the problem key(s):

What are the problems? Check all that apply:

e. Keys not well identified ___ f. Keys not well placed ___

g. Too many to learn & remember easily ___

h. Other ___ Describe problem briefly:

6. Are the amount of space and the layout of the work surfaces adequate for handling the paper records during page by page review of the files for image quality comparisons?

a. Yes ____ b. No ____ If "No," describe the problem:

7. Any other problems with quality control?

8. Any especially good features of the current QC station?

9. Any suggestions for improving quality control?

10. If you could change only one thing at the QC station, what would it be?

Figure E-3, Page 2 of 2

CMSR Search Batch
Group 5

INSTRUCTIONS: Read thoroughly before starting.

(a) Search to identify the correct answers for each of the 10 inquiries.

(b) As you work, fill in the CMSR FILE SEARCH form:

  (1) Fill in your name, the date, and the group #

  (2) Fill in the line for "Time started" when you begin work on the group and the "Time Completed" line when you complete the full group of 10

  (3) List your answers in the same 1 to 10 order as the inquiries are given below; if one file gives a cross reference to another file that has more complete information, list both in your answer.

INQUIRIES:

1. Query: Alfred Jackson

2. Query: Joseph Zeigler, a Confederate soldier

3. Query: John Hiett, old family Bible says he has blue eyes and brown hair and was a cavalry lieutenant

4. Query: Elihu Boggs, a blacksmith with the cavalry

5. Query: Thomas J Critchfield, a cavalry soldier captured by the Union army

6. Query: James Aden, a cavalry sergeant

7. Query: Lewis Sheppard, a blacksmith with the cavalry

8. Query: John Dudley, who died while a Yankee prisoner

9. Query: Asa Howell, a cavalry soldier

10. Query: John M Lincoln, a cavalry soldier

Figure E-4

CMSR File Search Form

Staff searcher: ______________________ Date: ____________

GROUP # _______________

Time started _______ Time Completed _______

Results of Search

Instruction: For each numbered query conduct the search. Check the right choice for the result of the search. Enter the fcn - file control number - that goes with the outcome of the search.

Query # 1

_______ no match

_______ exact match - fcn(s) # ______________________

____

_______ list of possible matches, but more information needed to select the exact one

# of files on list ___________

fcns of files on list ______________________

____

____

Query # 2

_______ no match

_______ exact match - fcn(s) # ______________________

____

_______ list of possible matches, but more information needed to select the exact one

# of files on list ___________

fcns of files on list ______________________

____

____

Figure E-5, First Page

Query # 9

_______ no match

_______ exact match - fcn(s) # ____________________

____

_______ list of possible matches, but more information needed to select the exact one

# of files on list __________

fcns of files on list ____________________

____

____

Query # 10

_______ no match

_______ exact match - fcn(s) # ____________________

____

_______ list of possible matches, but more information needed to select the exact one

# of files on list __________

fcns of files on list ____________________

____

____

****

THIS PART COMPLETED BY NSZ REVIEWER

Total time to complete batch: ____________

Number of errors: __________ % of errors: __________

Figure E-5, Last Page

Staff Reference Data Collection Form

Date of interview or of filling out this form _______________

Person interviewed or filling out this form ________________

Interviewer _____________

1. Please tell how easy or hard it was for you to learn how to use the ODISS staff reference station:

   a. How easy or hard was it to learn to operate the staff work station on a scale of 1 to 10 (1 = easiest & 10 = hardest)?

   ______

   b. What term most closely describes how easy or hard it was to learn to operate the staff work station? Check only one.

   Very easy ___

   Somewhat easy ____

   Average ______

   Somewhat difficult _______

   Very difficult _________

   c. What were the hardest things to learn in operating the station?

2. After you have learned to operate the station, how well overall does the ODISS staff reference station work on a scale of 1 to 10 (1 = lowest rating & 10 = highest rating)?

   ____

3. Is the writing/printing on the images of the CMSR files easy to read on the screen?

   a. Always easy __ b. Usually easy __ c. Often difficult __

Figure E-6, Page 1 of 3

4. When needed for better readability, are the original resolution, image zoom and rotation functions easy to learn and to use?

   a. Yes ___   b. No ___   If "No," describe problem:

5. Are the code tables easy to use in building an index search?

   a. Yes ___   b. No ___

   If your answer is "No," which tables are the problems?

   c. All ___   d. Regiment ___   e. Company ___   f. Rank ___

   What are the problems?

   g. Retrieval of a table too slow ___

   h. Scrolling within a table too slow ___

   i. Code numbers hard to remember ___

   j. Other ___   Describe the problem briefly:

6. Is shifting between index lists and file images easy?

   a. Yes ____   b. No ____   If "No," describe the problem:

7. Is printing copies easy? a. Yes ____   b. No ____

   If "No," please describe the problem:

8. Are the function keys easy to use? a. Yes ___   b. No ___

   If "No," which keys cause the problems?

   c. All ___   d. A specific key or keys ___   List the problem key(s):

Figure E-6, Page 2 of 3

What are the problems?

e. Keys not well identified ___ f. Keys not well placed ___

g. Too many to learn & remember easily ___

h. Other ___ Describe problem briefly:

9. Any other problems with staff reference?

10. Any especially good features of the staff workstation?

11. If you could change only one thing at the staff workstation, what would it be?

12. Any other suggestions for improving staff reference?

13. Compared with the current way of servicing CMSR records how do you rate the ODISS method on a scale of 1 to 10? ________
(1 = lowest rating & 10 = highest rating)

Please briefly explain your rating:

Figure E-6, Page 3 of 3

Public Reference Workstation Data Collection Form

Date of interview or of filling out this form ______________

Person interviewed or filling out this form ______________

Interviewer ____________________________

1. Learning to use the public reference system

   A. Are the written instructions on the screen adequate to teach you how to use the system?
   a1. Yes ____ a2. No _______

   Please explain briefly:

   B. Please tell how easy or hard it was for you to learn how to use the ODISS public reference station on a scale of 1 to 10 (1 = easiest & 10 = hardest)? ______

   b1. What term most closely describes how easy or hard it was to learn to operate the public work station? Check only one.

   Very easy ______ Somewhat easy _______
   Average _______
   Somewhat difficult _____ Very difficult _______

   b2. What were the hardest things to learn in operating the station?

2. Is the writing/printing on the images of the CMSR files easy to read on the screen?

   a. Always easy ____ b. Usually easy ____
   c. Often difficult ____

3. When needed for better readability, are the image zoom and rotation functions easy to learn and to use?

   a. Yes ___ b. No ___ If "No," describe problem:

Figure E-7, Page 1 of 2

4. Are the code tables easy to use in building an index search?

a. Yes ___ b. No ___

If your answer is "No," which tables are the problems?

c. All ___ d. Regiment ___ e. Company ___ f. Rank ___

What are the problems?

g. Retrieval of a table too slow ___

h. Scrolling within a table too slow ___

i. Code numbers hard to remember ___

j. Other ___ Describe the problem briefly:

5. Is shifting between index lists and file images easy?

a. Yes ____ b. No ____ If "No," describe the problem:

6. Is printing copies easy? a. Yes ____ b. No ____

If "No," please describe the problem:

7. Are the function keys easy to use? a. Yes ___ b. No ___

If "No," which keys cause the problems and what are the problems?

8. Any especially good features of the public workstation?

9. Any suggestions for improvements?

10. After you have learned to operate the station, how well over all does the ODISS public reference station work on a scale of 1 to 10 (1 = lowest rating & 10 = highest rating)? ______

Figure E-7, Page 2 of 2

APPENDIX F

RESEARCH TEST
IMPLEMENTATION CONSIDERATIONS

APPENDIX F. RESEARCH TEST IMPLEMENTATION CONSIDERATIONS

F.1 Rationale for Research Testing

The National Archives, in order to deal effectively with billions of paper records, needs to evaluate new information processing technologies continually. There are several ways to accomplish this, each with advantages and disadvantages. This can range from management studies to installation of a high volume production system. A basic marketplace survey, for example, involves vendor-supplied information. This data can be reviewed, and an analysis of the specifications can be informative. One major drawback is that specification sheets may not reflect actual performance under unique operating conditions.

A second approach, selected by NARA, is to install a system with all the capabilities of a larger system, but in a smaller configuration. A research test allows experimentation under operational conditions without the risk and expense of a major system procurement. The research system can be equipped with increased operational flexibility, which permits testing and analysis of alternative configurations.

A third approach is to install a large integrated system with all needed capabilities. This requires a greater initial capital outlay with inherently greater risk. This large-scale system approach requires a thorough understanding of all user requirements, and minimizes the opportunity for testing and analysis.

NARA determined that because of many reasons, not the least of which is the large volume of fragile, aged documents, a research test would be advantageous. Some benefits of a research test system are:

- The design concept is no longer strictly a piece of paper or abstract theory, but a mechanism that can actually perform functional work.
- The system can be used to test assumptions.
- Research systems can generally be installed in less time, and with less cost than full scale systems.
- User feedback can be obtained through actual system use.
- Research systems help show technological viability and can strengthen recommendations with greater certainty.
- Research systems help build communication bridges to users, and can help estimate productivity gains under live operational conditions.

A research test like ODISS permits system design and user requirements to evolve together prior to committing funds to a large system. ODISS has inherent design flexibility which supports alternative configurations, useful in testing different production and retrieval activities.

F.2 Role of the System Integrator

The federal government purchases millions of dollars worth of goods and services daily. Many are routine supply procurements, requiring minimal interaction between the requesting agency and the contractor. This was not the case with ODISS, which had technical specifications as part of a Request for Proposals. NARA's specifications outlined the government's expectations for ODISS functional and performance capabilities. The interested bidders reviewed the requirements, and provided their technical approaches to meet the ODISS goals and objectives.

ODISS's unique performance and operational requirements were not available in any readily available system. The technical specifications for each ODISS subsystem were demanding. The government looked to system integrators to propose state-of-the-art components tied into a cohesive system. Private industry corporations with strong system integration experience develop this integration capability over many years. This is accomplished by hiring, training, and retaining a professional staff with skill areas of: electrical engineering, systems analysis, software development, mechanical design, electro-optical engineering, documentation, and training knowledge.

Unisys provided such a staff in support of ODISS. Unisys's senior engineers were involved from the beginning, analyzing NARA requirements and formulating a system concept. This concept was examined in design reviews between the government and Unisys. Unisys engineers selected commercially available components from many different manufacturers, and provided the expertise needed to tie the disparate devices together into a cohesive system. The software engineers developed thousands of lines of code necessary to make ODISS operational. Due to ODISS complexity, the Unisys project team worked on separate subsystems under a configuration management plan. This required coordination of software and hardware in an ongoing effort, even though the basic design decisions were made early in the project.

The Unisys staff involvement did not end with system acceptance. Unisys provided on-site maintenance for one year following system acceptance, during which time the site engineer maintained contact with Unisys development staff for problem diagnosis and correction.

F.3 ODISS System Design Review Process

Government procurements for complex systems benefit from extensive communications between the system integrator and procuring agency personnel. This was accomplished in the ODISS project through a two-stage design review process: a system requirements review (SRR) and a critical design review (CDR). Although Unisys provided their basic design in the Technical Proposal, these meetings helped to ensure mutual understanding between Unisys and the government about project requirements.

The System Requirements Review was held shortly after contract award on October 28-29, 1986 in Camarillo, CA. This meeting was purposely held prior to Unisys making extensive progress in formal system development. It was an opportunity for Unisys to present their system concept before any significant hardware purchases or software development occurred. Preliminary equipment specifications presented at the SRR were useful for design and construction of the ODISS room. The basis of the SRR was presented in preliminary software development plans, work breakdown structures, program milestone plans, and equipment installation and facility site plans. Topics such as workflow processes, physical handling of

documents, display screen formats and menus, index modifications, and other basic criteria were covered in the two-day session.

A Critical Design Review was conducted at the National Archives in Washington, D.C. on December 16-17, 1986. It's purpose was to review Unisys's detailed design solutions, plans, and schedules to ensure that they satisfied contract requirements. This CDR addressed technical as well as contractual issues, and required more advanced software, hardware, and functional descriptions documentation. Attendees included Unisys project managers and engineering staff, NSZ ODISS staff, and NARA contracting officials. Unisys's engineers presented each major system function, using block diagrams and flowcharts for illustrative purposes. The Unisys plan for information workflow was also described. Planned hardware items were discussed, and modifications to off-the-shelf components were highlighted. Information concerning video display screen menus was presented for comments. Topics such as Unisys's hardware substitutions and changes to system documentation were discussed. Another topic was the acceptance and testing of the entire system and its individual components. Mutual understanding of this key point was significant since payments were tied to performance milestones.

NARA's technical staff concluded that the design met the technical requirements, and that Unisys successfully completed the Critical Design Review.

F.4 Factory Acceptance and On-Site Testing

The ODISS IFB specified several levels of testing during the life of the contract. Factory testing gave Unisys an opportunity to demonstrate that the system was fully integrated prior to shipment. The on-site tests were designed to verify system operations, throughput capabilities, and system reliability. A factory acceptance test (FAT) plan was provided by Unisys under the terms of the ODISS contract. This document was a test guide and data collection source. The Unisys-designed test plan incorporated a systematic approach to verifying system development and implementation. Testing verified functional and performance capabilities of individual components and the integrated ODISS system.

Two factory tests were conducted prior to shipment, the first was held during February 1-5, 1988. A NARA team visited the Unisys Corporation in Camarillo, CA, where the test plan specified verification of equipment components as stand-alone devices and demonstration as a fully integrated system . Although many functions were verified as operational during this first FAT, system integration and connectivity were not working up to expectations. This first FAT was determined to be not successful.

Project schedules and plans had to be revised to compensate for a second FAT test. Unisys was provided with time to correct the deficiencies. A second FAT demonstration was held during May 16-20, 1988. The NARA team returned to Camarillo after notification that all problems were corrected and thoroughly tested. The test team decided to redo all of the test plan criteria because of the major deficiencies noted during the first FAT. In general, the second FAT was a better indication of the systems' capabilities, although not every requirement was demonstrated. Unisys's engineers were working on software development up to the last minute, and under these conditions new problems had been introduced. The government team gave Unisys an opportunity to work on the problems, but some required more extensive corrective efforts than were possible under the testing environment. The major problems again centered on the system's inability to work as a cohesive entity, and its

inability to demonstrate simultaneous operations. This second FAT test was also determined to be unsuccessful.

The system was shipped to NARA after Unisys staff corrected the FAT test deficiencies. After installation in the main Archives building, an on-site 30-day system reliability test began on July 25, 1988. NARA staff recorded all system problems, which were factored into up-time calculations. An 87% system up-time was recorded after thirty days of operations.

Unisys decided to correct all the recorded deficiencies, and worked on-site for over seven weeks which required Unisys staff to be relocated from Camarillo to Washington, D.C. A second 30 day test was completed on December 6, 1988 at a reliability level of 92%. ODISS now met all performance requirements, and remaining contract payments to Unisys were approved.

F.5 ODISS Facility Design and Construction

Available floorspace at the National Archives is in short supply, and acquiring sufficient ODISS space required the involvement of the NARA administrative and facilities department. The first step defined the minimum space requirements. Several potential sites were reviewed, and subsequently not accepted due to insufficient square footage or inadequate room configuration. The final ODISS site was previously used as office space, necessitating a move by the existing occupants. In order to get the ODISS room built, several concurrent activities were undertaken. After ODISS contract award, Unisys provided equipment layout and signal cable installation plans, power receptacle requirements, and electrical power and floor loading specifications. A detailed set of architectural drawings were prepared based on the Unisys supplied information. The drawings went through the typical review process of several drafts, until the final drawing set was approved by NARA and Unisys. The drawings covered room electrical, heating/cooling, construction, lighting, and fire suppression systems.

Based on these drawings, an ODISS room construction contract was awarded. This contract was monitored by the General Services Administration. Due to the planned one-year ODISS system delivery schedule, the room project received a priority status. The room design required a changeover from office space to computer room architecture. The entire area was dismantled, and reconstructed with raised flooring and dropped ceilings. The space was divided into one large production area with a small separated workspace for Unisys's site technician. The room was completed and the workstation furniture and other support equipment were installed prior to ODISS system delivery. A security alarm system was installed on both ODISS entryways and connected to the NARA guard station.

F.5.1 Computer Room Environment

The ODISS room was designed as a computer type facility, with a six-inch raised floor system for cable management. Due to the quantity of energy-consuming equipment planned and the attendant BTU heat generation, some type of auxiliary air cooling was mandatory. An under-floor plenum process cooling system was considered but not approved. Since NARA has a year-round chilled water system, a series of fan coil units were installed. The fan coil units circulate chilled water based on thermostat settings, and a three-speed fan can be set to the desired velocity. Five fan coil units were installed in the ODISS room, and proved capable of adequate cooling.

Fan coils produce cool drafts which affected ODISS system operators. Continual fan setting adjustments upset the ODISS room temperature stability. It was difficult to maintain a consistent temperature suitable for the operations staff and the computer equipment. A second disadvantage of fan coil units is their dependence on building chilled water for operation. Several building chilled water outages required ODISS to halt operations. This prevented damage to the computer processors from elevated room heat. The existing building ventilation was not sufficient to cool the ODISS room under a chilled water outage.

Computer room florescent lighting diffusers were specified and installed in the room. The lights were controlled by three separate wall-mounted switches. In practice, the lighting system was incompatible with high resolution display monitor usage. The ambient room light levels lowered screen contrast, requiring several of the light banks to be turned off. An improved system would include adjustable light dimmer controls.

F.5.2 Fire Safety and Control Systems

The fire suppression system for the ODISS room used a complex, automatically controlled process. A main control panel[114] monitored ceiling and under-floor smoke detectors. The fire suppression system itself used Halon gas as a primary defense, with a water nozzle system as a backup. In case of fire or smoke, the controller would turn off the room power, and sound an alarm. The highly audible alarm would sound for thirty seconds, followed by Halon release. The system was tested and passed a GSA fire safety engineer's test. If a fire is not extinguished by the gas, then the water backup system will take over. It is preferable that the Halon extinguishes the fire rather than water, due to the potential damage to the computer equipment by the latter. The room has a ventilation system for removing residual Halon gas. In case of a false alarm, the Halon system can be halted by an emergency hold button. Releasing Halon gas without cause is economically unsound, and environmentally unwise.

F.5.3 Electrical and Signal Cable Installation

The ODISS electrical layout was based on Unisys-provided equipment power specifications, acquired in sufficient time to allow room design to proceed. The electrical requirements affected equipment installation and room cooling capacities.

Careful attention was directed to electrical plugs, receptacles, and connectors. Unisys specified the equipment requirements, and the room construction contractor was responsible for the needed components. Electrical receptacles were installed under the raised flooring, attached to the concrete sub-flooring. Electrical wiring was installed in accordance with the GSA construction drawings.

GSA's room contractor installed a Unisys-provided power conditioner. This device stabilized incoming building power and served as a spike/surge filter to avoid damaging sensitive ODISS equipment. To avoid contaminated current backflow, the high speed scanner was wired into a dedicated circuit. An emergency OFF button, capable of instantly shutting off electricity to all ODISS room equipment, was also installed.

[114] Refer to Figure H-8 in Appendix H.

The ODISS room received raised flooring to accommodate the communication signal cables. All of the devices within the room, as well as two NARA remote workstations are interconnected using signal cables. Unisys provided all of the required cabling materials, which were specified as teflon coated, and plenum rated. NARA electrician staff installed the cables between the ODISS room and the Microfilm Reading Room and area 7E1 for the public and staff remote terminals respectively. Unisys's engineers made the final hard-wired connections between these stations and the main ODISS system. The cables within the ODISS room were custom-prepared in Camarillo as part of the factory system development process. These cables were delivered and installed by Unisys during the on-site system installation. The cables and device receptacles were uniquely labelled to ensure proper connections.

F.6 Equipment Floorplan Design

The ODISS equipment floorplan design (Figure F-1) was an evolutionary process. Initial workflow and workstation design concepts were used to pre-identify square footage space requirements within NARA. Following ODISS contract award, discussions between the government and Unisys project personnel were held concerning facility specifications and equipment installation requirements. Workflow processes were an integral part of the final layout configuration, with an emphasis on paper and electronic image pathways. The need to store and transport cartloads of documents around the ODISS workstations during ODISS operations, and the overall room shape influenced workstation placement. NARA staff created sketches of planned workstation designs. These scaled room drawings allowed design analysis to be done on paper prior to committing to a final configuration. This information was provided to the GSA architectural contractor for use in their part of the facility design process. Figure F-2 shows the design of the workstations which were located outside the ODISS room in other parts of the building.

F.7 Ergonomic Workstation Furniture Specification

Since ODISS was a new project with extensive keyboarding tasks, computer room furniture was selected. Off-the-shelf furniture was obtained from Steelcase Corporation. One custom-made Steelcase cabinet was ordered for the public station's printing hardware. Operator functions at each of the ODISS workstations were analyzed using industrial engineering techniques. This was done to ensure that the furniture would support rather than impede the daily operations. A local Steelcase representative worked closely with NARA staff to assure the best match between workstation design and operator activities. Computer tables with various keyboard supports, cut-outs, and mechanically adjustable work surfaces were installed. EckAdams provided the computer room chairs, and an adjustable stool for the high speed scanner was obtained. Other support items, such as staff lockers, a technician's workbench, printer stands, storage cabinets, and acoustical divider panels were installed. The furniture was an integral ODISS system component, rather than being an add-on afterthought. This integration provided benefits during the production phases through increased operator productivity. Any future system should apply the same analytical techniques to operators workstation designs.

ODISS Floor Plan

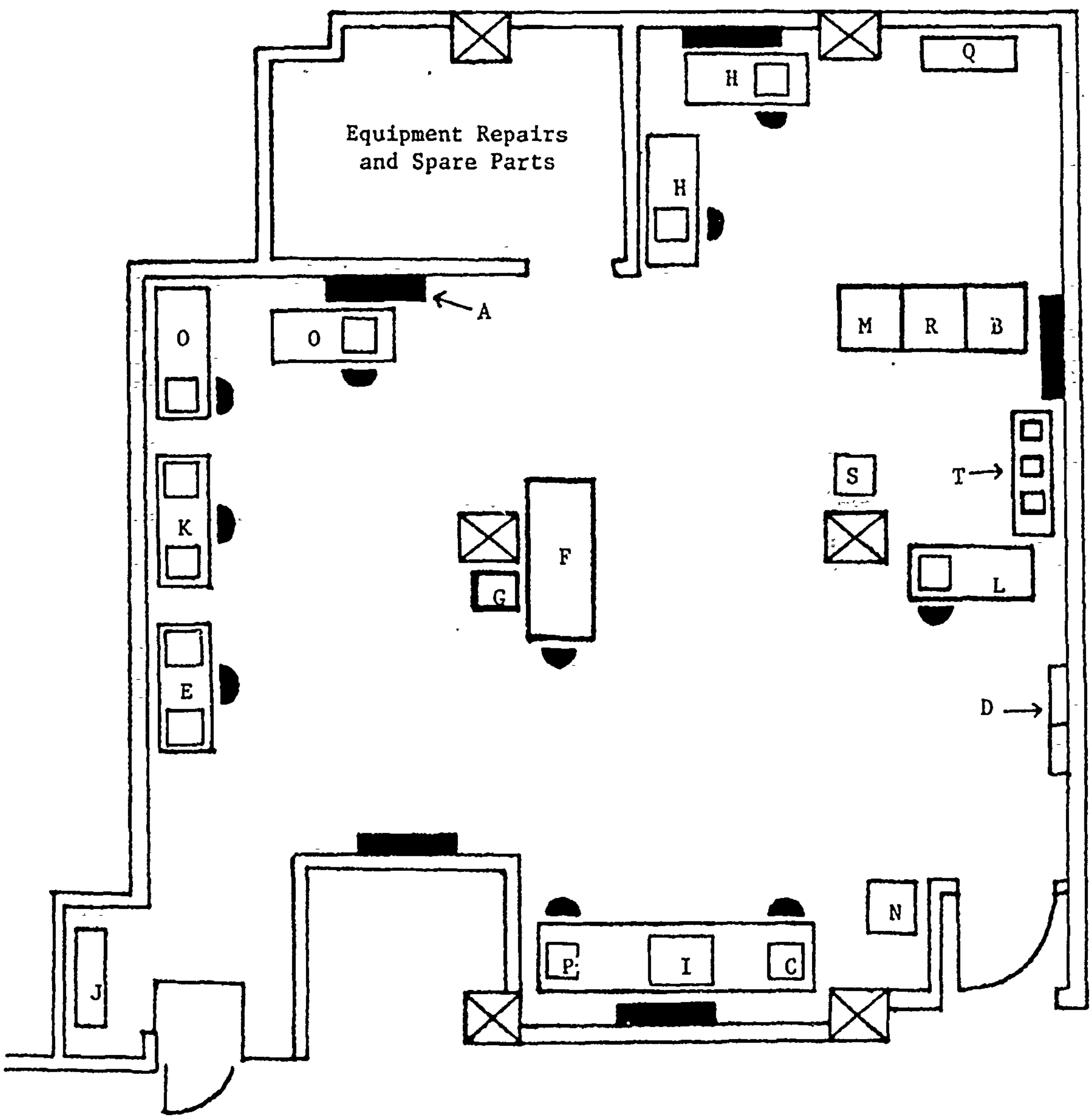


Figure F-1, Page 1 of 2

Key to ODISS Floor Plan

Match the following to the floor plan schematic drawing:

A= Air Handling and Conditioning Units (5 each)

B= Core Computer Hardware Cabinet

C= Demonstration Retrieval Workstation

D= Electrical and Halon Control Panels

E= Gray Scale Scanner Subsystem

F= High Speed Scanner

G= High Speed Scanner Electronic Controller

H= Index Workstation

I= Laser Printer and Print Controller

J= Lockers and Coat Rack

K= Low Speed Scanner Subsystem

L= Multiformat Microform Scanner

M= Optical Disk Jukebox

N= Power Line Conditioner

O= Quality Control Workstation

P= Staff Reference Workstation

Q= Supply Storage Cabinet

R= System Manager Hardware Cabinet

S= System Manager Printer

T= System Manager Computer Control Terminals

Figure F-1, Page 2 of 2

Workstation Designs

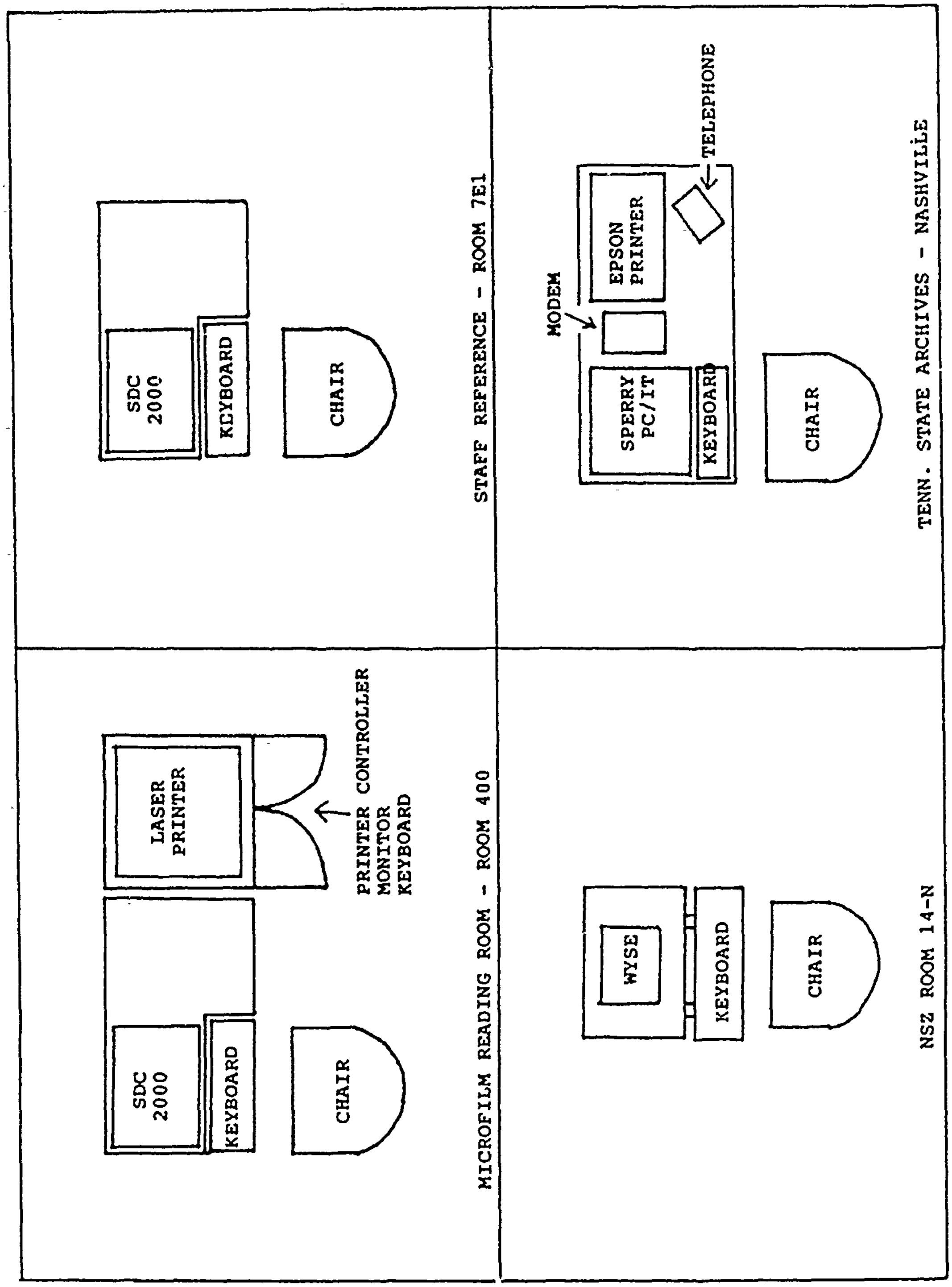


Figure F-2

F.8 Production Staff and User Training

Automated, high technology, image processing environments require employees with advanced skills and abilities. ODISS, for example, required skills not normally found in entry level governmental positions, and sufficient staff to fill the required positions were not readily available. Therefore, NSZ requested staff from NN with certain basic skills and experience, and then provided extensive training as part of the ODISS test.

NSZ defined the skills needed by staff to run the installed ODISS system. There was a need for workstation operators and a system manger to administer and coordinate overall system operations. It was agreed that NN would provide and supervise the ODISS operations personnel. NSZ would direct the research test and interface with the contractor about fulfilling technical requirements.

NN initially provided a staff of eight operators and a general supervisor. One of the eight was selected to perform most of the system manager duties, and others were trained to provide back-up assistance to this system manager. Unisys provided formal classroom training following system installation. The ODISS system documentation served as textbooks and reference manuals. The training was held for ODISS operators, system managers, and NN staff members who would be using the system. The training included an introduction to the ODISS system with training in each production step. Scanner and workstation training was provided, including troubleshooting and minor problem solving. This was followed by hands-on guidance during subsequent weeks. ODISS operators were cross-trained in the various station functions so that work assignment rotations, and coverage for absent staff were possible. During the course of the project there was some staff turnover, and at times staffing fell below the initial level. Replacement ODISS operators received training on each station from the system manager and co-workers.

System users were trained in basic operations as part of the ODISS data collection efforts for this report. Selected users were tested and queried about system performance and user interface design. Extensive efforts were expended to make ODISS utilization as self-explanatory as possible. This effort was useful in reducing user training, as instructional information is provided in screen displays.

F.9 System Documentation

The ODISS IFB specified several levels of ODISS contract documentation requirements, divided into four main categories: technical, system oriented, project administration, and training materials.

Technical manuals were complete descriptions of ODISS system hardware, software, and system functional capabilities. The descriptions were written for a wide audience, and included areas such as theory of operations, detailed operational modes, system level block diagrams and flow charts, special procedures for operations and repairs, and manufacturer specification sheets and interface requirements.

System manuals covered the use, operation, and maintenance of the research test system. These manuals provided general system component descriptions, operational and backup modes, detailed instructions on use of terminals and workstations, operational procedures for workflow and system control, special tools and trouble shooting guidelines, and other descriptive and illustrative materials deemed necessary to support the ODISS operation.

Project Administration included plans, schedules, and design documents; letter progress reports providing updated communications regarding project progress; equipment installation and site preparation plans useful for system installation and electrical power needs; supply item specifications for required consumables; and factory and on-site test plans.

Training documentation, materials, and schedules for operator and management instruction courses were provided according to an approved schedule.

APPENDIX G

NARA MICROGRAPHICS PROGRAM

APPENDIX G. NARA MICROGRAPHICS PROGRAM

G.1 Technology Overview

Microfilming originated in the nineteenth century, with early uses primarily related to military surveillance applications. Commercial applications gradually arose out of the need to reduce document storage space and improve information retrievals. The banking industry embraced microfilm for recording large volumes of customer's financial transactions. Although the early systems primarily used roll microfilms, a variety of film formats eventually evolved. In order to keep pace with competitive industries, micrographics has progressed beyond basic data storage functions. Computers and microprocessors are integrated into innovative approaches to solving information retrieval problems. Today the industry offers image management systems which meet a wide range of user needs. Micrographic systems range from single user installations, up to large-scale, around-the-clock production operations employing hundreds of technicians under the control of computer process control systems.

A potential micrographics user is faced with a series of decisions revolving around project costs, microformat selection, in-house versus service bureau production, search and retrieval techniques, user equipment, facility modifications, equipment maintenance, and data security. To aid the user community, the information processing industry has professional system integrators and management consultants offering comprehensive technical guidance.

Micrographic system implementations are ideally based on a series of studies conducted prior to procuring production hardware. Accurately defining the existing records systems problems, and determining current system operations are important initial development steps. These studies should be followed by a detailed system design, which specifies factors such as equipment descriptions, system performance criteria, and projected overall costs. These steps are ideally followed by planning efforts for the project implementation, followed by system installation and integration. Follow-up reviews of the installed system are important to determine if the system is operating to the designed specifications.

Micrographics systems typically require specialized hardware to perform the conversion operations. It is highly recommended that prior to any hardware procurement a thorough analysis of system requirements be performed. Micrographic equipment consists of cameras, film processors, quality control inspection, duplication equipment, and related accessories. Skilled operations staff are also required to perform the needed conversion activities. System operations documentation is valuable in providing staff guidance and in solving production problems.

Micrographics involves image capture using some type of camera recording device. Depending on input document characteristics and conversion time requirements, cameras range from single-sheet, hand-fed table-top cameras, up to fully mechanized document transport systems. For microfilms which require it, film development is usually accomplished with automated roll film processors. The developed films are quality-inspected, with duplicates produced for user reference.

Some common microforms are roll films, microfiche, aperture cards, and microjackets, while others exist for unique applications. Roll films offer relatively low production costs, compact storage, and inherent file integrity. NARA has made extensive use of roll films in its

microfilming program. Various roll film indexing schemes, and computer-assisted retrieval (CAR) systems are available to aid information retrievals.

Microfiche, aperture cards, and microfilm jackets are commonly referred to as unitized microforms. These formats are ideal for storing related information, but usually require somewhat more complex production equipment. Microform jackets involve the cutting and insertion of roll films into the microjacket film channels. Unitized microforms are easily reproduced, and are suitable for automated retrieval equipment. Microfiche and microfilm jackets typically contain a title area for data identification, and aperture cards can be key punched for machine processing.

Unique formats have been developed to address specific user needs. A few examples include: high reduction microforms containing thousands of images on small film areas, strip-up type microfiche systems, and image retrieval systems using large format roll films. Specialized formats usually require custom production equipment and techniques.

An important records management aspect is file updating. Some microforms are especially designed with an update capability, while others require physical alterations such as manual splicing of roll films. Updatability is useful for active records systems which require additions to the image database.

Micrographics and computers have been successfully linked through computer output microforms (COM). COM recorders have historically served as alternatives to computer line printers. Digital imaging systems are prime candidates for microform output equipment. A film recorder with raster capabilities integrated into a digital imaging system can provide archival-quality microform output. Computers are also actively used in CAR systems, storing images on film and using computers for database and system management functions.

Information retrievals require viewers to enlarge microimages to human-readable size. Typical systems require duplicate microforms, with the master films securely stored under archival conditions. Viewers are available for all common formats and reduction ratios, many with automated features. Hardcopy output from microimages is handled by viewer printers. The micrographics industry offers several printing technologies; selection should be based on quality images and low print costs.

G.2 Camera Area Equipment and Operations

G.2.1 Equipment

Microform cameras are precision devices, designed to capture images at greatly reduced sizes while retaining fine line details and content. Over the years, NARA has acquired nineteen cameras:

- Eleven Kodak MRD-2 Series Planetary Filmers
- Five Terminal Data Corporation Multi-format DocuMate I Cameras with computerized microfiche titling
- Two Terminal Data Corporation DocuMate II Cameras with microfiche titling
- One SMA 35mm flatbed camera

NARA has a large volume of fragile, aged, bound, and other difficult-to-handle documents which are best suited for hand-feeding. Kodak MRD planetary microfilmers require hand placement of individual documents. A planetary camera is one in which the document and film are stationary during film exposure. This design yields optimum image sharpness, and minimizes potential document handling damage. NARA's MRD cameras require manual exposure adjustments to compensate for the wide range of document characteristics. The MRD's output is 16mm or 35mm film, with operator selectable reduction ratios. The majority of NARA microfilming is with 35mm films at 14X reduction.

Terminal Data Corporation (TDC) multi-format cameras offer 16mm through 105mm output, with quality optics for image sharpness. NARA owns two different TDC microfilmers: Documate I cameras for single-sheet, manual hand-feeding; and Documate II's with mechanized belt-type document transports. Documate I's are used extensively by NARA for microfiche production. The Documate II's transport belts change direction after the front side is filmed, followed by automatic reverse-side filming. The only NARA records currently processed by the Documate II cameras are relatively sturdy index card stock documents. NARA's TDC cameras are equipped with computerized microfiche titling systems.

The SMA flatbed camera is an engineering drawing device for 35mm films. This camera accommodates large-sized originals, and it also is used for color microfilming. Color films require out-of-house developing by a local film processing company.

NARA has acquired various document hold-down and book cradle devices to accommodate the diverse records holdings. Physical document characteristics directly impact microfilming throughput rates since any special handling requirements reduce the time available for microfilming throughput. Some very difficult-to-handle documents require careful placement under a glass platen, which holds the document flat and improves image sharpness.

G.2.2 Staffing

NARA's camera area currently has seven full time employees. At various times, NARA had up to nineteen staff members occupied in microfilming activities. The current operations staff includes personnel knowledgeable in camera equipment operations, film processing, and silver film duplication. The staff grades include:

- Four General Schedule grade four (GS-4) operators
- Two General Schedule grade five (GS-5) operators
- One General Schedule grade seven (GS-7) supervisor

The camera staff is on NARA's production standards program. The camera room is located on the east side of the 19th floor of the National Archives Building. Two rooms contain the varied collection of cameras and accessories. Shelves and tables for holding the records are also located in the microfilming area.

G.2.3 Production Costs

Microfilm camera production costs include expenses for personnel, equipment, supply, and handling costs. Document characteristics often mandate which camera is used and the daily throughput rate per camera. NARA's fees for microfilming are $0.36 for each 16mm camera

frame and $0.37 for each 35mm camera negative frame produced.[115] These prices include all production steps including pulling the documents and mailing the completed products. For the most part, NARA's existing camera equipment is of sufficiently advanced age to be fully depreciated. This reduces the cost per page for major equipment items. The major cost items are direct personnel costs for archival handling of the documents, camera operations, film processing, quality inspection, and mailing.

A NARA employee cash awards program is available to microfilm camera operators who routinely exceed NARA's established production rates. For example, production from 150% to 183% over base rates will result in a $300 cash award per quarter. Operators who can produce rates of 184% over base will receive a $400 bonus per quarter.

G.3 Processing Equipment and Operations

G.3.1 Equipment

The exposed microforms are forwarded to NARA's film processing laboratory located in Room B-5, which contains the following equipment:

- One Kodak Prostar II processor for 16mm and 35mm films
- One Kodak Versamat processor for 16mm to 105mm films

NARA processing conditions are monitored to ensure consistent development and precise film specifications. The developed films are tested twice weekly for residual thiosulfate, an important factor in archival quality microforms intended for permanent retention. The research and testing lab located in Room B-3 conducts the archival film tests, the results of which are logged for future reference. Films which fail the test are rewashed and then retested. The film processing lab is equipped with water filtration systems required to remove suspended particles from the incoming city water supply.

The film processing area also has film duplication equipment:

- One Extek Silver Roll-to-Roll Film Printer
- One Extek 5101 Cut Microfiche Printer
- One 3M Diazo Microfiche Duplicator System
- One CX Microfiche Cutter

This equipment creates duplicate roll films and microfiche. The majority of 16mm and 35mm duplication is performed by NNPS-D at NARA's South Pickett Street annex in Alexandria, Virginia. The Extek 5101 accepts cut microfiche, and produces roll microfiche as output. The Extek roll-to-roll printer produces copies of roll films and microfiche at high throughput speeds. The silver prints require film processing, with 105mm film rolls cut into individual microfiche with the CX microprocessor-controlled cutter.

[115] NARA fee schedule for microfilm services; 1989-90.

G.3.2 Staffing

NARA's film processing station requires personnel familiar with the equipment, processing chemistry, and film production specifications. The position requires monitoring the processing solutions, operating the film processing equipment, and creating film duplicates. A low-volume diazo duplicate microfiche system is also available. The technician inspects films, and maintains log books for archival film test results.

G.4 Quality Control Equipment and Operations

G.4.1 Equipment

NARA's film inspection program conforms with Federal Property Management Regulations (FPMR) requirements. The film inspection equipment located in Room B-5 includes:

- Digital X-Rite Densitometer
- Microscope
- TDC Viewer Mate for microfiche inspection

A densitometer measures exposure and development on the processed microfilm. A microscope helps evaluate test targets included on each film roll. The TDC viewer displays either an entire microfiche, or a single frame on a large viewing screen.

G.4.2 Staffing

A film processing technician monitors not only one specific roll of film, but also the overall camera and processing conditions. More detailed inspection occurs after the films are returned to the camera area. The record's custodial unit is responsible for any additional inspection performed.

G.5 Duplication Equipment and Operations

G.5.1 Equipment

The majority of high-volume microfilm duplication is currently performed at NARA's Pickett Street Annex, which has the following equipment:

- Two Extek roll-to-roll duplicators
- Four B&H Carleson duplicators (16mm and 35mm films)
- Allen deep tank film processors
- Microfilm inspection equipment

G.5.2 Staffing

NARA's microform duplication section has seven full time employees:

- Six Wage Grade sevens (WG-7)
- One Wage Grade eight supervisor (WS-8)

The duplication staff is assigned to the Pickett Street Annex, where the duplication equipment and print masters are maintained. The staff operates film inspection stations to

check for exposure, film development, master-to-copy tracking, and image resolution (sharpness). Approved duplicates are packaged to fulfill customer orders.

G.5.3 Production Costs

Production costs vary based on duplicate film type. To complete all the necessary production steps including accepting the request, pulling the correct print master, printing and processing, and delivery of the duplicate, the per-foot costs are $0.32 and $0.34 for 16mm and 35mm direct negative duplicates respectively. Positive polarity print per-foot costs are $0.31 (16mm), and $0.33 (35mm).[116]

G.6 Future Plans

NARA's microform production and user equipment was acquired over many years. Because of NARA's high volume utilization, replacement hardware is required to replace worn out, existing devices. Micrographic operations function best when located in facilities specifically designed to meet production requirements. In NARA's main building and the Pickett Street Annex, extensive alterations were needed to install the power, water, drainage, cold storage, lighting, and related support services. This becomes especially difficult when the location was not originally designed for the tasks, or when major renovations are needed.

The new Archives II building planned for the University of Maryland's College Park Campus includes an integrated micrographics capability. This facility will offer larger, expanded spaces to house the various production operations. Space needed to combine NARA's various microfilming sections will be available. The NNPS and NNPD groups will be located in a carefully planned spaces. Individual camera booths will allow ongoing production without interference from neighboring stations. Film processing will have adequate incoming water and power supplies, and silver recovery systems will be installed to minimize environmental impacts. The NNPD duplication area will have adequate print master storage, and additional silver duplicate processing equipment can be installed. This arrangement will allow combining all 16mm to 105mm duplication under one management and logistic control center.

[116] *Ibid*

APPENDIX H

PHOTOGRAPHS OF ODISS EQUIPMENT

APPENDIX H. PHOTOGRAPHS OF ODISS EQUIPMENT

High Speed Scanner

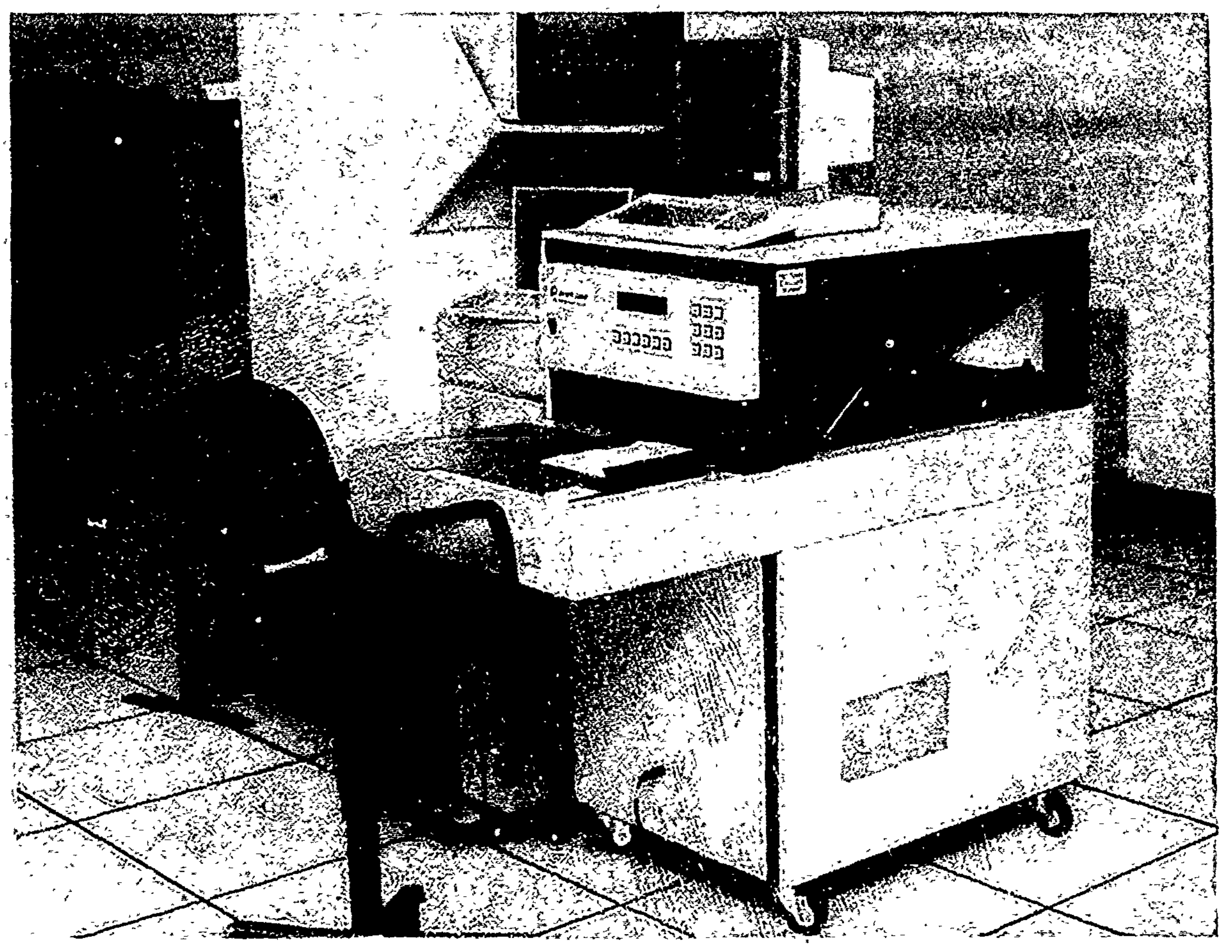


Figure H-1

Low Speed Scanner Station

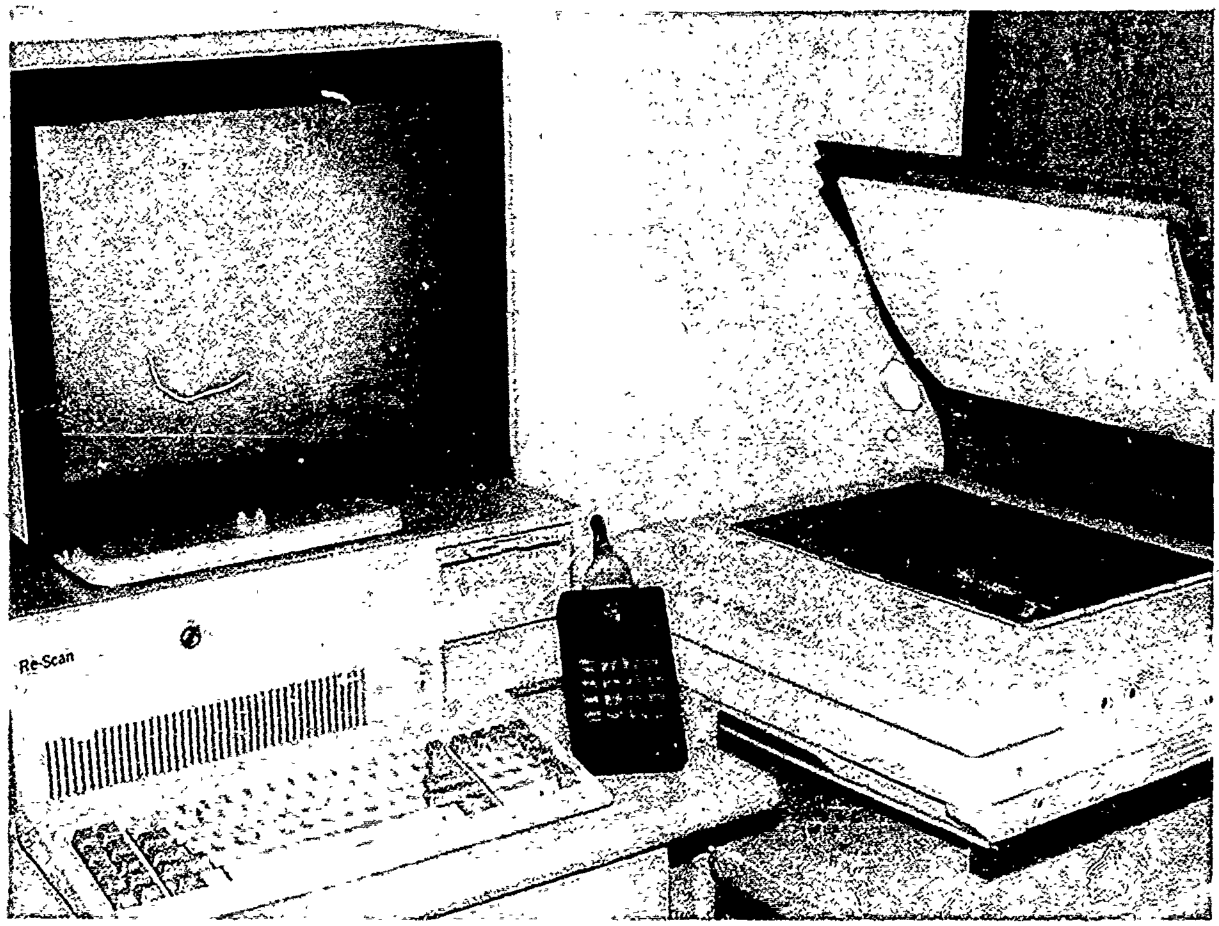


Figure H-2

Microfilm Scanner Station

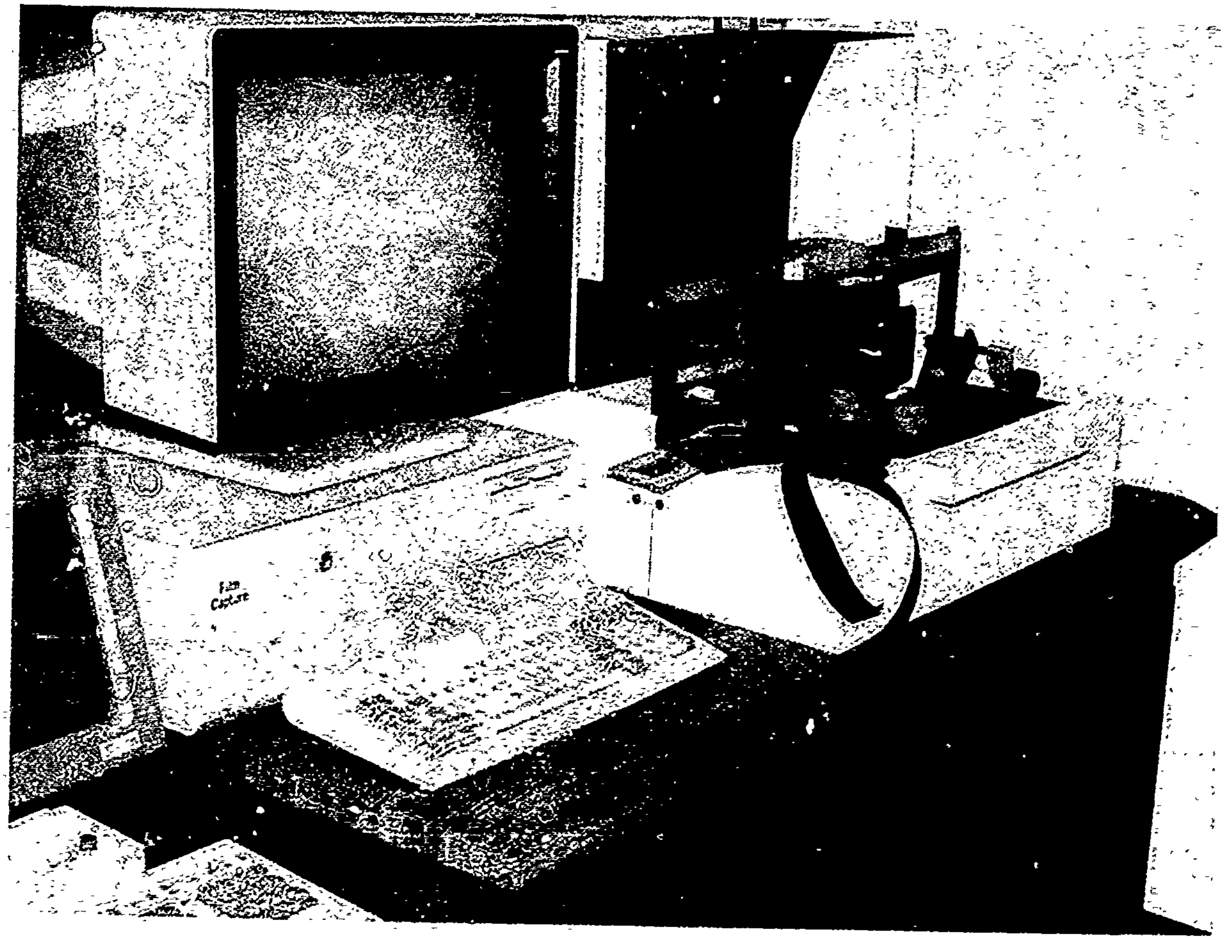


Figure H-3

Index and Quality Control Stations

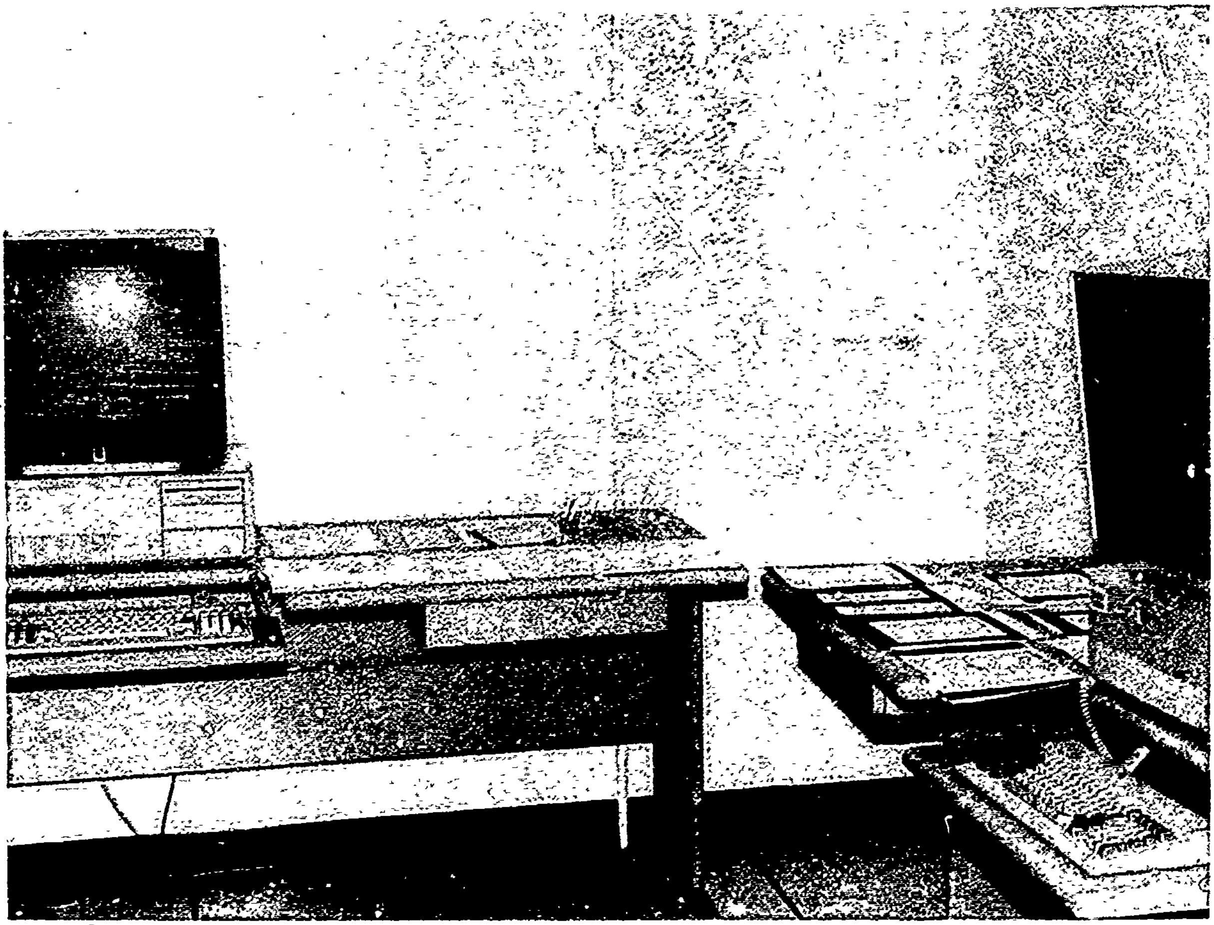


Figure H-4

Optical Disk Jukebox

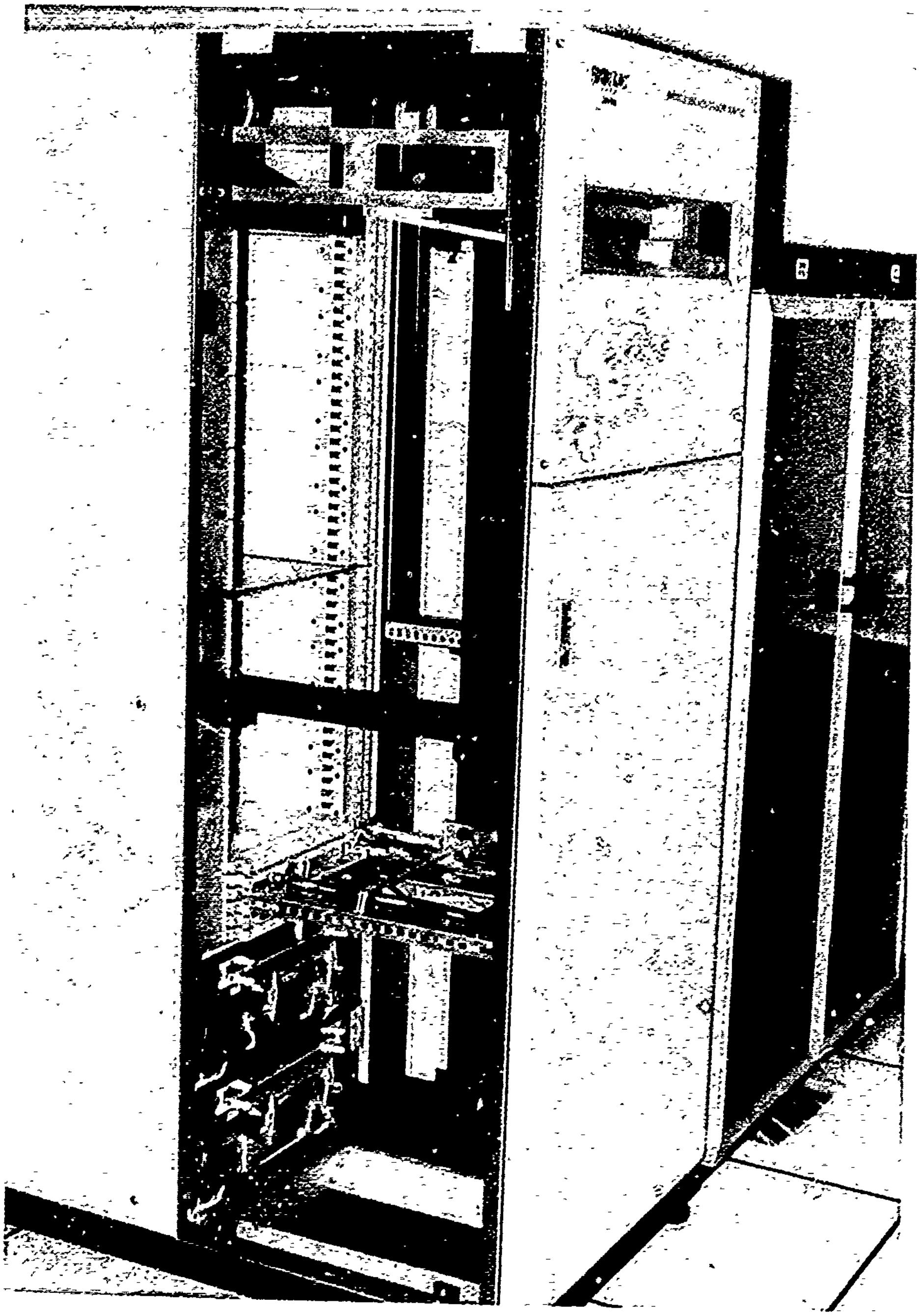


Figure H-5

Retrieval and Printing Stations

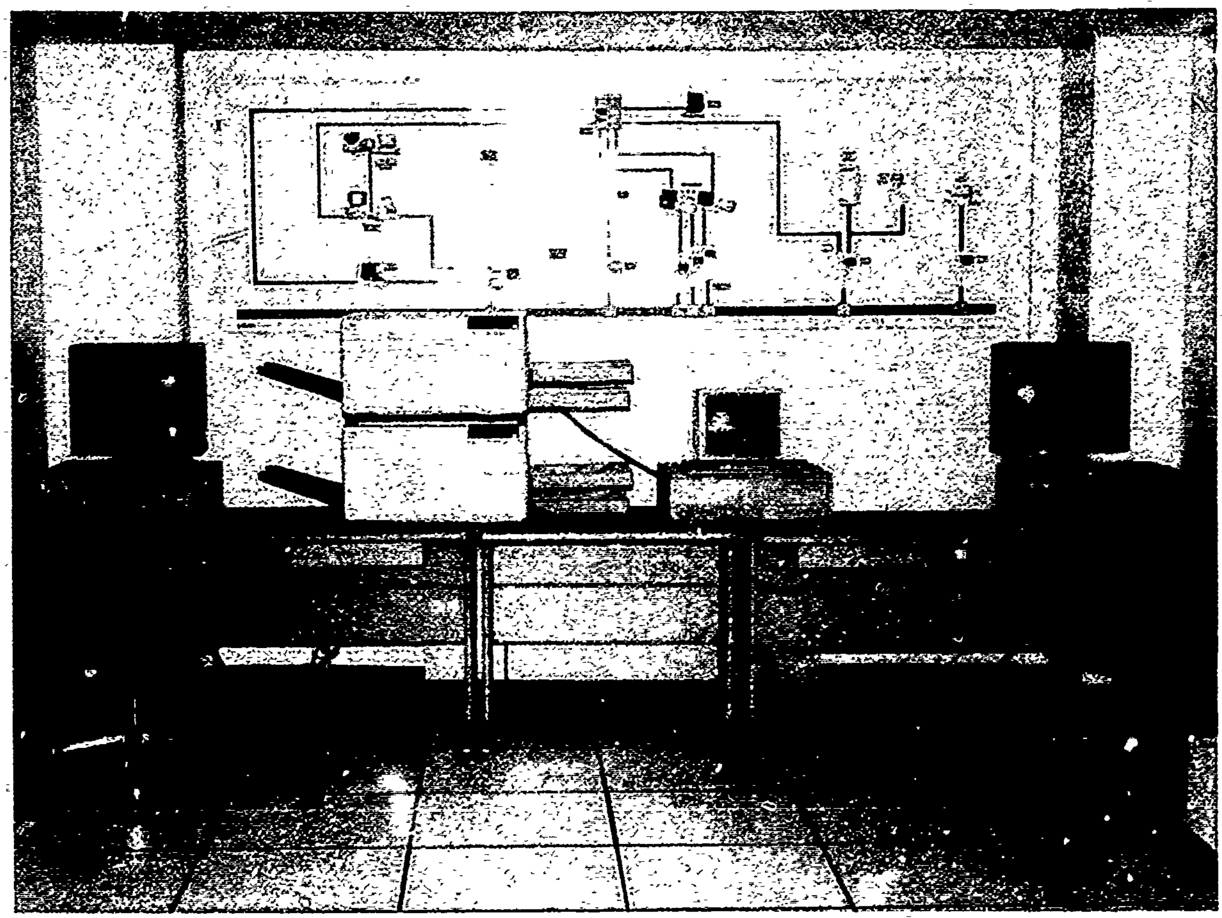


Figure H-6

System Manager Station

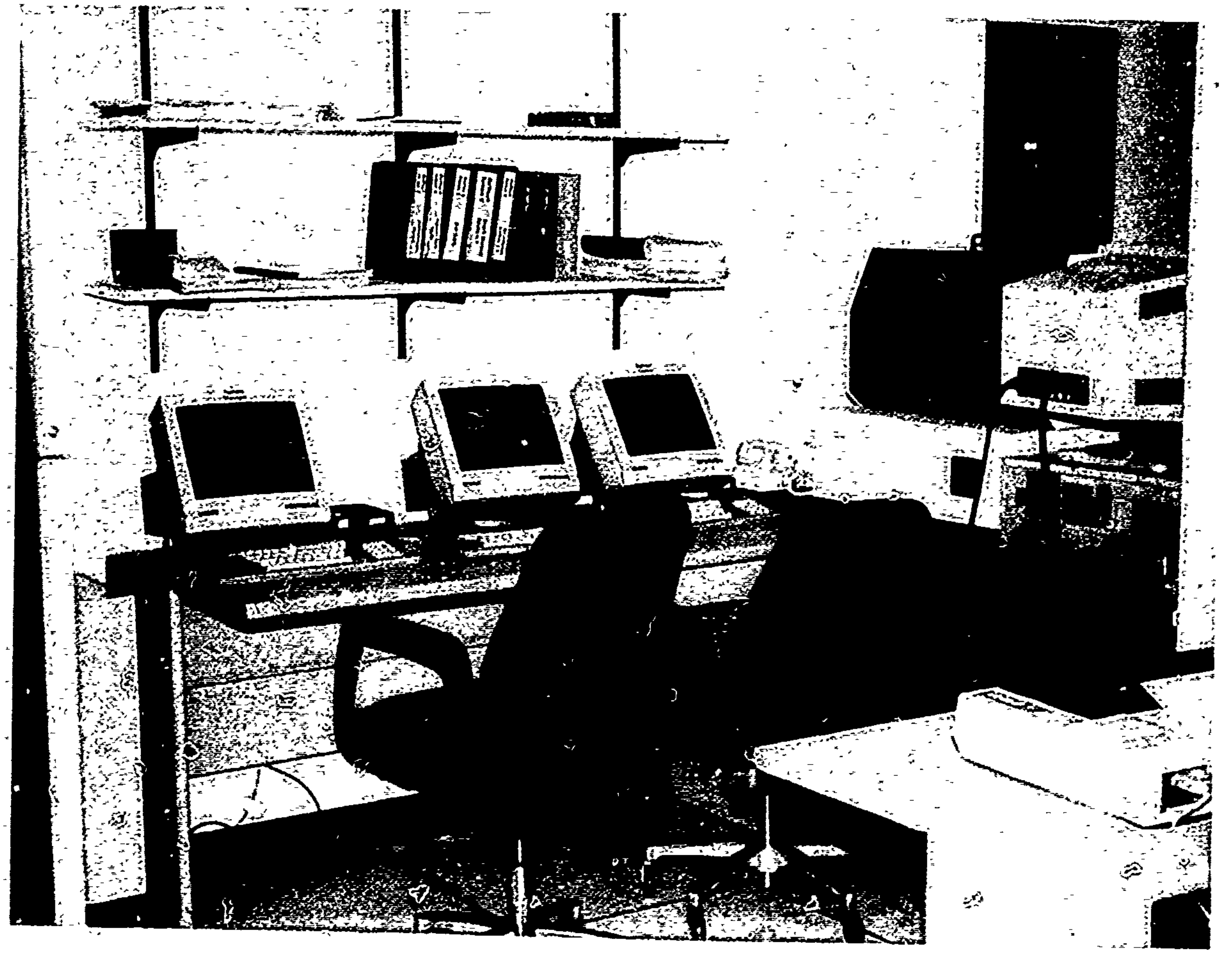


Figure H-7

Halon Fire Control Panel

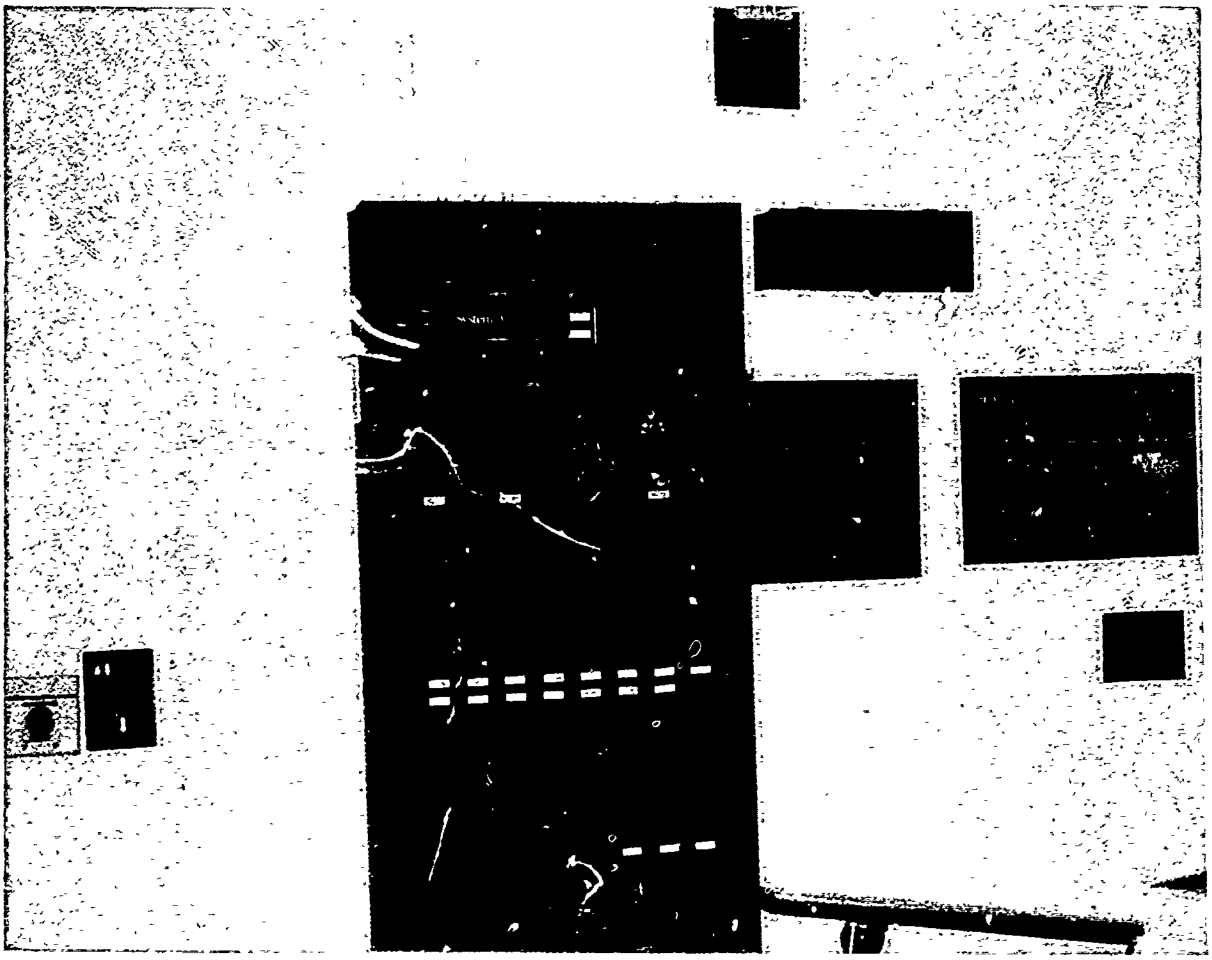


Figure H-8

APPENDIX I

GLOSSARY OF TERMS

APPENDIX I. GLOSSARY OF TERMS

This glossary lists and defines some of the technical terminology used in the body of the report. Following the definition is a listing of pages on which the terms are used.

Algorithm: A formula for solving a problem; a set of steps in a very specific order, such as a mathematical formula or the instructions in a computer program.
4, 9, 10, 11, 12, 69, 78, 108, 109, 112, 153, 172, 224, 227, 233

Analog Data: Data or information which is stored or transmitted using electronic signals which vary in amplitude and/or frequency.
58, 164, 165, 179, 180, 189, 195, 203, 210, 222, 223, 225

Analog Videodisc: Optical storage in which information is carried on a signal that continually varies according to the range of image intensity and frequency. Each disc side can store up to 54,000 separate photographic images, or up to one hour of full-motion video.
179, 189, 195

Bandwidth: Data communications term to describe the amount of frequency variance necessary to carry the digital information signal.
19

Beta-test: Term typically applied to installing equipment at an operational site for purposes of performance testing and analysis prior to mass production and marketing activities.
10, 58, 112

Binarization: Term applied to one-bit image processing which produces only pure blacks and whites from numerous intermediate levels of gray. This process results in the lowest storage requirements for captured images. Binarization is generally accomplished through a process called thresholding.
116

Binary Scanner: A one-bit scanner system which records only the black and white digital information.
57, 58, 111

Bit: A single <u>bi</u>nary digi<u>t</u> (0 or 1) in the binary number system. Groups of [usually eight] bits make up storage units called characters or bytes.
10, 59, 108, 109, 113, 167, 171, 172, 173, 179, 193, 202, 211, 213, 222, 223, 225, 228, 235, 243, 264, 277

Blip Marks: Small geometric shapes, frequently squares or dots, recorded along one edge of a roll of [micro]film, which are used by automated film transports to locate and position the film to a particular image frame.
11, 12, 120, 302, 311

Booting: Process of initializing computer operations once electric power is applied to the system equipment.
146, 153, 275, 278

Byte: A unit of computer storage holding the equivalent of a single character.
58, 109, 115, 116, 119, 122, 123, 124, 171, 228, 229, 264

Byte Storage: Amount of digital storage expressed in its equivalent character or byte capacities.
115, 116, 119, 122-124

CAV: Constant Angular Velocity refers to one technique used to record information on the optical disks. CAV optical media store information in concentric tracks. CAV disks spin at a constant rate of speed, so the spacing of data stored on inner tracks is more compact than that for data stored in tracks nearer the outer edge of the disk. CAV disks provide faster access rates than CLV disks, but sacrifice storage capacity per square inch. (ODISS uses CAV disks.)
82, 126, 243

CCD: Charge-Couple Device is the electronic scanner component which senses changes in reflected light intensities and converts these changes into an analog electrical signal.
58, 95, 149, 151, 165, 166, 210, 211, 213, 214, 222, 225, 230, 231

CCD Sensitivity: The range of color sensitivity and ability to recognize various subtle color differences by the CCD sensor.
95, 149, 151

CCITT: Acronym for Consultative Committee on International Telephones and Telegraphy.
172, 211

CLV: Constant Linear Velocity describes a data storage technique for optical media in which data are stored in one continuous helical (spiral) track. The disk is rotated at a variable speed so that data are stored along the track at a constant spacing irrespective of its distance from the center of the disk. CLV discs can store more data but have slower data access rates than CAV disks.

Compressed File: A file whose data has been compressed to a smaller size by application of one or more data compression algorithms. In ODISS, image files are compressed in order to reduce storage requirements and to facilitate faster data transmission between system components.
119

Constant Threshold: An image enhancement process by which each pixel is resolved to either pure black or pure while depending on whether its shade lies above or below some arbitrary level of gray (i.e., the threshold).
9, 93, 109, 149, 167, 234

Continuous Tone: An image containing various shades of gray (as opposed to only pure black and white), requiring halftoning and gray scaling techniques for best image reproduction.
116

Contrast: Term referring to the degree of difference between the lighter and darker areas of an image, with high contrast images consisting of tones nearer the extremes of blacks and whites and with few intermediate shades of gray.
8-10, 13, 20, 42, 47, 66, 67, 73, 91, 93, 94, 109, 111, 112, 121, 124, 125, 153, 166, 167, 233, 344

Contrast Stretch: An image enhancement algorithm in which lower and upper percentage "saturation" parameters are selected. All pixel intensities lying between these two limits are "stretched" toward their respective [back and white] extremities, resulting in increased visual image contrast.
109, 228, 234

Convolution Filter: A specific image enhancement technique or algorithm, in which the operator selects the filter size and weights, and image scaling parameters.
211, 213, 234

Data Backup: To create a duplicate copy for security or disaster recovery purposes. ODISS created backup copies of data stored on both optical and magnetic disks.
195

Data Compression: Algorithmic techniques by which redundant digital data streams are reduced to much smaller sizes, resulting in lower storage and data transmission requirements. ODISS used Group III, one-dimensional CCITT standard compression techniques.
171, 211, 235

Density (film): Photographic term referring to the amount of light transmitted through a [micro]film image, as measured with a precision inspection device.
42, 124

Density (ODDD): With respect to optical digital data disks, refers to the storage compaction techniques utilized in recording information on the disk.
9, 10, 12, 21, 96, 116, 117, 121, 152, 165, 166, 175, 193, 240

Digital Data: Data or information which is stored or transmitted as a sequence of discrete, off-and-on electronic signals.
2, 24, 147, 164, 179, 195, 202, 210, 222, 225, 322

Digital Image: An electronic data file consisting of digital data, that when reconstructed either on a display screen or hardcopy print, appears as a facsimile of the original document.
2, 3, 6, 7, 9, 11, 13, 14, 19, 21-24, 26-30, 54, 55, 57, 60, 61, 73, 74, 84, 92, 105, 109, 110, 117, 120, 121, 125, 139, 140, 142, 148, 149, 151, 152, 160, 163-167, 169-174, 178, 179, 201-203, 206, 210, 214, 231, 239, 246, 284, 290-293, 302, 305, 307-309, 315-319

Dots Per Inch: DPI or dots per inch is a method of defining image resolution or definition. DPI is linked to pixel sizes, with smaller pixels [and greater DPI] yielding increased image definition. ODISS scanned images at 200, 300, and 400 dots per inch.
3, 9-12, 57, 58, 60, 94, 96, 112, 115, 116, 119, 122, 165, 172, 175, 178, 233, 252

Dynamic Threshold: A sophisticated image enhancement technique in which each pixel on the page is individually thresholded based upon the shading of its surrounding pixels.
121, 167

Edge Detection: Image enhancement algorithms in which the system highlights the boundary edges of image data giving a visual sharpness or distinction to the edges of shapes in the image.
234

Enhancement: The process of using electronic algorithms to "clean up" or intensify a digital image in order to improve its contrast or legibility.
4, 9-12, 26-28, 57, 58, 61, 69, 78, 79, 81, 93, 94, 103, 108-112, 115, 117, 121, 122, 124, 150, 153, 154, 164, 166, 167, 173, 206, 222, 223, 225, 226, 228-233, 235, 239, 308

Generation: Refers to the status of an image in relation to the original document or microform master. An image captured from the original document is a "first generation" copy. A copy made from a first generation copy is a second generation copy, and so forth.
12, 13, 24, 124, 133, 149, 202

Gigabyte: One gigabyte is equivalent to one billion (1,000,000,000) computer encoded characters. Also refers to the amount of computer storage necessary to hold one billion characters.
62, 126, 193, 307, 315

Gray Scale: Refers to the capture of data representing the various shades of gray existing in the typical document images The amount of storage required to hold an image is related to the number of levels of gray captured and retained. The greater the number of gray shades kept, the greater is the storage requirement. (Gray scaled images nearly always require greater storage volumes than binarized images.)
8-10, 57, 58, 108, 109, 112, 116, 120, 166, 167, 169, 222, 225, 230, 234, 235

Halftone: Reprographic process in which various screens are employed during printing plate/ink processes to improve continuous tone photograph images. Digital imaging systems employ electronic techniques to simulate the process, creating varying sizes of black dots.
58, 94, 109, 111, 120, 234, 235

Hardware Process: Digital scanner image enhancement capability which uses hard-wired circuitry and components built into the scanner's electronics.
57, 94, 108, 109, 115, 153, 222

High Speed Scanner: A scanner capable of high production rates, usually using a transport system which moves the documents past light sources and CCD arrays which are permanently mounted in fixed positions. Some models (such as the ODISS scanner) are capable of scanning both sides of a document on one pass through the scanner.
6, 7, 9, 19-20, 28, 57, 58, 60, 68, 70, 74, 91, 93-96, 98, 105, 107, 112, 115-117, 119, 120, 122, 124, 152, 153, 155, 160, 166, 171, 210, 211, 213-215, 217, 223, 231, 235, 239, 252, 269, 283-289, 307, 308, 344

Host Computer: The primary or main controlling computer system.
58, 230, 243

Image Capture: Term relating to the acquisition and recording of a [facsimile] image on some type of storage media.
6-8, 22, 23, 26, 28, 33, 39, 54, 90, 92, 108, 111, 151, 153, 169, 235, 309, 315, 352

Image Workstation: A primary user reference tool in digital imaging systems, typically containing a high resolution screen capable of displaying a document image, and a keyboard for entering user commands. A printer may also be included.
174, 206, 260, 308

Index: Descriptive information associated with a file that enables a requestor to identify the file and retrieve it from the storage medium.
2, 4, 7, 15-19, 22-24, 26, 27, 39, 45, 47, 49, 54-56, 59-63, 83, 84, 90, 92, 98, 100-105, 108, 121, 133, 135, 136, 134-142, 145, 147-149, 152, 164, 169, 170, 173, 206, 236, 239, 240, 246, 252, 271, 277, 291, 297, 302, 317

Index Code Tables: Tables that define valid code equivalents used in index fields to stand for specific items of alphanumeric information. Usage of code values in indexes reduces storage requirements and generally improves access speeds.
56, 59-61, 76, 83, 99, 100, 136, 141, 149, 154, 237, 252, 267

Jacket: The envelope in which CMSR files typically are stored.
6, 16, 37, 54, 59, 60, 66, 67, 89, 92, 93, 95, 96, 120, 122, 124, 236, 237, 240, 328

Jukebox: Descriptive term applied to optical disk storage systems which utilize robotic devices containing shelves and automated picking mechanisms to store and retrieve multiple disks, thus providing rapid, automatic digital image delivery.
20, 55, 62, 63, 82, 83, 125, 135, 146, 193, 201, 202, 206, 210, 243, 245, 246, 277, 278, 307-309, 315

Laser Printer: A printer commonly used in electronic imaging systems, these non-impact devices utilize laser beams to create a temporary image on a photosensitive material. This latent image is developed by applying toner particles, which are subsequently transferred and permanently fused to create the paper print.
11, 73, 85, 115, 121, 125, 139, 173, 178, 206, 252, 260, 262, 307

LED: Acronym for light emitting diode, a device which when energized yields visible light. LED's are frequently used in computer systems as indicators of equipment function or status.
112, 174, 213

Lens Filter: An optical component consisting of either glass or plastic sheet material, typically installed on a lens to improve color sensitivity of the image capture system.
72, 95, 116, 152

Low-Contrast: A generic term referring to aged, faded documents which have faint image characteristics. These documents often have a variety of handwritten inks and paper stock colors, requiring careful application of thresholding and other image enhancement techniques.
10, 111, 112

Noise: Extraneous pixels typically appearing in the background of digital images which may detract from document image legibility.
10, 112, 115, 119, 121

OCR: Optical character recognition is a technology which translates the graphical representation of the document provided by the raster map to a character-based representation expressed in a character codeset.
26, 170, 307

ODDD: Acronym for optical digital data disk.
82

PC/IT: Unisys's version of an 80286-based, IBM-compatible, personal microcomputer.
214, 217, 223, 225, 237, 260, 262

Pixel: Abbreviation of "picture element", one of large number of small dots that collectively comprise a digital image. Usually referred to as number per inch, such as 200 pixels per inch (or dots per inch). At 200 pixels per inch, a one-inch square would contain 200 x 200 for a total of 40,000 pixels.
3, 4, 28, 57, 58, 109, 115, 151, 165, 167, 171, 172, 175, 178, 213, 214, 217, 222, 228, 233, 234, 235

Reduction Ratio: Term used frequently in micrographics to express the size ratio of the original document to the microimage, such as 24:1.
120, 121, 165, 352, 354

Reflectance Surface: The surface which lies behind a document being scanned on a scanner. The cover on a platen scanner generally has a white reflective surface.
10

Resolution: See dots per inch.
11, 57-59, 85, 91, 93, 94, 115, 116, 119, 121, 122, 152, 165, 174, 175, 178, 222, 225, 231, 237, 252, 264, 308

Scanner: The hardware component in a digital imaging system which converts the original to an electronic image. ODISS used high and low speed document scanners, and a multiformat film scanner to capture image data.
3, 6, 7, 9-11, 19-20, 28, 45, 54, 55, 57-60, 68, 73, 90-98, 103, 105, 107, 108, 111, 112, 115-117, 119-125, 152-153, 155, 160, 164-167, 171, 178, 206, 210, 211, 213-215, 217, 222, 223, 225, 230-232, 235, 239, 250, 269, 283-288, 307, 344

Scrolling: Image display technique in which a user can pan horizontally across, or scroll vertically up and down an image using either a mouse or cursor keys.
61, 76, 140, 175, 245, 251, 325, 335, 338

SCSI: Acronym for Small Computer System Interface, a standard controller interface widely used with optical disk drives.
61, 62, 243, 245, 246, 266

Search: Any use of the [CMSR] index to identify and/or retrieve file images.
4, 15-18, 21, 22, 24, 29, 45, 49, 55, 61, 84, 135-139, 148, 149, 173, 240, 246, 252, 262, 269, 273, 301, 335

Software Process: Image enhancement system employing computer software (vice hardware) to perform image processing.
10, 20, 57, 58, 109, 111, 153, 222

System Manager: Term referring to the ODISS system module or workstation used for central control of the system; also refers to the ODISS staff member who serves as the system production manager.
32, 55, 56, 60-63, 70, 78, 79, 92, 98, 100, 107, 108, 111, 122, 125, 126, 144-148, 150, 154, 156, 214, 217, 223, 237, 239, 245, 246, 260, 266, 267, 269-273, 278, 279, 349

Terabyte: One trillion (1,000,000,000,000) computer-encoded characters of storage. Also refers to the amount of computer storage necessary to hold one trillion characters.
195

Thresholding: Enhancement algorithm by which each pixel is resolved from a shade of gray to either pure black or pure white depending on whether the shade of gray lies above or below the threshold value. Thresholding can be applied to all or a portion of an image using a "constant" or a "dynamic" process.
9, 11, 73, 90, 92-94, 109, 111, 112, 121, 149-152, 153, 167, 212, 227, 231, 234, 235

Threshold Setting: Scanner operator action using push buttons or keyboard function keys to specify the desired type and level of image processing threshold to be used to binarize a document image.
90, 92, 111, 121, 149, 150, 167, 231, 234

Tiling: A method of scanning large size documents, in which subsections are scanned separately, and electronically joined by the computer system to create a digitized image representing the whole document. Each subsection of the document is called a tile.
165

Transport Sensor: Optical or vacuum detectors designed to "sense" the presence of a document, looking for skewing or other document feeder problems.
7, 95-98

UNIX: A multi-user, multi-tasking operating system, developed by AT&T Bell Labs, that can be used on all classes of computers from microcomputers to mainframes.
57, 145, 146, 239, 260, 266-269, 272, 278

Wait Time: Term referring to the interval between the periods when an operator is able to key-enter data or commands at the terminal. During the "wait" period, the system is performing certain actions that temporarily disable the terminal.
20, 32, 74, 75, 93, 100-102, 104, 106-108, 154, 155, 160, 162, 173

Work Time: Term referring to the interval of time needed for an operator to perform a specific function.
75, 102, 106, 107, 155, 160

WORM Disk: An acronym for "Write-Once, Read-Many" times optical disks which can store (write) user data and can be accessed (read) when needed. The data recorded on WORM disks is considered permanent, in that the disks are not erasable or reusable like conventional magnetic media.
4, 13, 19, 55, 133, 193, 195, 197, 201, 203, 206, 307, 308

☆ U.S. GOVERNMENT PRINTING OFFICE: 1991 — 292 - 334/91104